用户指南

NeuralTools

Microsoft® Excel 神经网络插件

第 5.7 版
2010 年 7 月

Palisade Corporation
798 Cascadilla St.
Ithaca, NY 14850
USA
+1-607-277-8000
+1-607-277-8001（传真）
http://www.palisade.com（网站）
sales@palisade.com（电子邮件）

版权通告

版权所有 © 2010, Palisade Corporation。

商标确认

Microsoft、Excel 和 Windows 是 Microsoft Corporation 的注册商标。
IBM 是 International Business Machines, Inc. 的注册商标。
Palisade、TopRank、BestFit 和 RISKview 是 Palisade Corporation 的注册商标。

欢迎使用 Excel NeuralTools

欢迎使用

NeuralTools 为行业标准数据分析和模型建立工具 Microsoft Excel 提供了一套功能强大的建模工具！NeuralTools 是一款 Microsoft Excel 神经网络插件，允许您在 Excel 工作表中分析数据，并在熟悉的 Microsoft Office 环境中工作。通过将功能强大的数据管理程序与最先进的神经网络算法相结合，NeuralTools 为您提供以下两个领域的最佳性能：Microsoft Office 便于使用和报告的功能，以及神经网络强大且准确的预测功能。

在您熟悉的环境中工作

如果您熟悉 Excel，您就会熟悉 NeuralTools！NeuralTools 的工作方式与 Excel 完全相同，工具栏、菜单和自定义工作表函数全部都在 Excel 中。与独立神经网络软件不同，由于操作方法与 Excel 中的完全相同，因此 NeuralTools 不需要花费很长的学习时间和使用前培训费用。您的数据和变量都在 Excel 电子表格中。您可以使用标准 Excel 公式以及 Excel 数据排序表和数据透视表进行计算。您的分析报表和图表用标准的 Excel 格式显示，并可以利用所有 Excel 的内建格式化功能。

NeuralTools 分析

神经网络能够学习数据中的复杂关系。通过模仿大脑的功能，神经网络能够发现数据的模式，然后在获得新数据时进行推断预测。使用神经网络解决的问题可以分为大类：

- **分类问题：**尝试确定某一未知项目属于哪一类别的问题。其中包括医疗诊断和预测信用付款能力。
- **数字问题：**需要预测具体数字结果的问题。其中包括股票价格预测和预测未来某段时间的销售水平。

神经网络用于各类应用，其中包括：股票市场预测、信用和贷款风险分配、信用欺诈探测、销售预测、一般业务预测、投资风险、医疗诊断、科学领域研究和控制系统。

NeuralTools 包括最新神经网络算法，对分类问题（在 NeuralTools 中称为范畴预测）和数字问题进行最佳预测。

NeuralTools 数据管理

NeuralTools 在 Excel 中提供完善的数据集和变量管理程序，与随 StatTools（Palisade 为 Excel 提供的统计插件）提供的数据集和变量管理程序类似。您可以直接从您 Excel 中的数据定义任意数量的数据集，并且每个数据集都带有需要分析的变量。NeuralTools 以智能化方式访问您的数据块，建议变量名称、类型以及数据位置。您的数据集和变量可以位于不同的工作簿和工作表中，允许您按照自己认为适当的方式组织数据。然后，对参考您的变量的神经网络进行训练，而不是反复在 Excel 中重新选择数据。此外，NeuralTools 中变量的大小并不受限于 Excel 工作表中的一列数据 — 您可以将跨度高达 255 份工作表的同一列用于一个变量！

NeuralTools 报表

Excel 非常适用于编写报表和图表，而 NeuralTools 可以最大限度地利用这一功能。NeuralTools 利用 Excel 格式的图表，用户可以方便地自定义新的颜色和字体，并添加文字。报表标题、数字格式和文字的更改方法与任何标准 Excel 工作表中的操作一样。从 NeuralTools 报表中直接将表格和图表拖放至您在其他应用程序中的文档内。

NeuralTools 正式版还包括在将新数据输入 Excel 工作表的同时计算预测值的实时预测功能。如同其他 Excel 重新计算一样，实时计算会自动进行。

数据访问与共享

Excel 具有出色的数据输入功能，因此将您的现有数据输入 NeuralTools 也十分方便！使用标准 Excel 功能读入来自 Microsoft SQL Server、Oracle、Microsoft Access 或任何其他 ODBC 兼容数据库的数据。从文本文件或其他应用程序载入数据 — 如果可以将数据读入 Excel，就可以通过 NeuralTools 使用读入的数据！

NeuralTools 将所有结果和数据保存在 Excel 工作表中。与任何其他 Excel 文件一样，您可以将您的 NeuralTools 结果和网络发送给任何地方的同事。共享极为简便！

NeuralTools 专业版和正式版

NeuralTools 提供两种版本 — 专业版和正式版。这两种版本的差异如下：

- NeuralTools 专业版的数据集不得超过 1000 个案例，而 NeuralTools 正式版最多可以支持包括 16,777,216 个案例的数据集。
- 在将新数据输入您的 Excel 工作表的同时计算预测值的实时预测功能仅在 NeuralTools 正式版中提供。如同其他 Excel 重新计算一样，实时计算会自动进行。

目录

第一章：入门指南 vi

简介 3

检查您的软件包 3
软件包内装物品 3
本版本简介 3
在您的操作环境中使用 4
如果需要获得帮助 4
NeuralTools 系统要求 6

安装说明 7

一般安装说明 7
设置 NeuralTools 图标或快捷方式 7
DecisionTools Suite 8

软件激活 9

第二章： NeuralTools 概述 13

概述 15

为什么要使用神经网络？ 15
NeuralTools 和神经网络 15
NeuralTools 菜单和工具栏 16
数据集和数据集管理器 16
训练神经网络 18
检验网络 22
预测 23
NeuralTools 报表和图表 25
NeuralTools 实用工具 26
将 NeuralTools 与 StatTools、Solver 和 Evolver 配合使用 26

第三章：NeuralTools 参考指南 27

简介 29

参考：NeuralTools 图标 31

NeuralTools 工具栏 31

参考：NeuralTools 菜单命令 33

简介 33

对话框中的图标 34

命令参考 35

数据集管理器命令 35
训练命令 40
检验命令 51
预测命令 58

实用工具 61

应用程序设置命令 61
神经网络管理器命令 64
缺失数据实用工具命令 66

神经网络的更多信息 69

神经网络基本知识 69

神经网络与统计方法 69
神经网络的结构 69
数值预测和范畴预测 70
训练网络 70
神经网络的计算机处理 70
神经网络的类型 70

多层前馈网络 71

MLF 结构 71
训练 MLF 网络 72

广义回归神经网络和概率神经网络 75

广义回归神经网络 75
概率神经网络 77

MLF 网络与 PN/GRN 网络的比较 79

输入项变换 81

推荐读物 83

索引 85

第一章：入门指南

简介 3

检查您的软件包 3
软件包内装物品 3
本版本简介 3
在您的操作环境中使用 4
如果需要获得帮助 4
NeuralTools 系统要求 6

安装说明 7

一般安装说明 7
设置 NeuralTools 图标或快捷方式 7

简介

本简介描述 NeuralTools 软件包包含的内容，并向您说明如何安装 NeuralTools 以及如何将 NeuralTools 随附在 Microsoft Excel 2000 (Windows 2000) 或更高版本的副本中。

检查您的软件包

您的 NeuralTools 软件包应当包括：

NeuralTools 或 DecisionTools Suite CD-ROM，其中包括：

- *NeuralTools 程序*
- *NeuralTools 教程*
- *《NeuralTools 用户指南》（本手册，PDF 格式）*

NeuralTools 许可协议

如果您的软件包不完整，请致电您的 NeuralTools 经销商或供应商，或者拨打电话 +1-607-277-8000 直接与 Palisade Corporation 联系。

软件包内装物品

NeuralTools 可以单独购买，也可以与 DecisionTools Suite 专业版和正式版一起购买。NeuralTools CD-ROM 包含 NeuralTools Excel 插件、几个 NeuralTools 示例，以及完整索引的 NeuralTools 在线帮助系统。DecisionTools Suite 专业版和正式版包括以上所有项目，以及附加应用程序。

本版本简介

本 NeuralTools 版本可以作为 Microsoft Excel 2000 或更高版本的 32 位程序安装。

在您的操作环境中使用

本《用户指南》假设您掌握了 Windows 操作系统和 Excel 的一般知识。特别是以下知识：

- *熟悉自己的计算机和鼠标的用法。*
- *熟悉图标、单击、双击、菜单、窗口、命令和目标等术语。*
- *理解目录结构和文件命名之类的基本概念。*

如果需要获得帮助

我们向拥有有效维护计划的所有 NeuralTools 注册用户免费提供技术支持，或按照每次事件收取服务费。要确保您是 NeuralTools 的注册用户，**请在以下网站注册：http://www.palisade.com/support/register.asp。**

如果通过电话与我们联系，请准备好您的序列号和《用户指南》。如果您坐在计算机前并作好准备，这将有助于我们为您提供更好的技术支持。

打电话之前

在与技术支持部联系之前，请检查以下核对清单：

- *是否已经查阅了在线帮助？*
- *是否已经查阅本《用户指南》，并查看了联机多媒体教程？*
- *是否已经阅读 README.WRI 文件？该文件包含可能未纳入手册的有关*NeuralTools *的最新信息。*
- *是否是能够始终重复出现的问题？是否是能够在另一台计算机或不同型号的计算机上重复出现的问题？*
- *是否已经在互联网上查阅了我们的网站？我们的网址为：**http://www.palisade.com**。我们的网站还在“技术支持”部分中提供最新常见问题解答（可搜索的技术支持问答数据库）和* NeuralTools *补丁。我们建议您定期访问我们的网站，了解有关* NeuralTools *以及其他 Palisade 软件的最新信息。*

与 Palisade 联系

Palisade Corporation 欢迎您提出有关 NeuralTools 的问题、评论意见或建议。请用以下任何一种方法与我们的技术支持部人员联系：

- *请发送电子邮件至 support@palisade.com。*
- *请在美国东部时间上午 9 点至下午 5 点（星期一至星期五）致电 +1-607-277-8000，并按照提示接通技术支持部。*
- *请发送传真至 +1-607-277-8001。*
- *请寄邮件至：*

  Technical Support
  Palisade Corporation
  798 Cascadilla St.
  Ithaca, NY 14850
  USA

如果您希望与 Palisade 欧洲分部联系：

- *请发送电子邮件至 support@palisade-europe.com。*
- *请致电 +44 1895 425050（英国）。*
- *请发送传真至 +44 1895 425051（英国）。*
- *请寄邮件至：*

  Palisade Europe
  31 The Green
  West Drayton
  Middlesex
  UB7 7PN
  United Kingdom

如果您希望与 Palisade 亚太分部联系：

- *请发送电子邮件至 support@palisade.com.au。*
- *请致电 +61 2 9929 9799（澳大利亚）。*
- *请发送传真至 +61 2 9954 3882（澳大利亚）。*
- *请寄邮件至：*

  Palisade Asia-Pacific Pty Limited
  Suite 101, Level 1
  8 Cliff Street
  Milsons Point NSW 2061
  Australia

无论您用何种方法与我们联系，请提供产品名称、版本和序列号。请在 Excel 的 NeuralTools 菜单上选择“帮助”命令，查找确切的版本。

不为 NeuralTools 学生版提供技术支持。如果需要获得帮助，我们建议您采取以下替代方法：

- *向您的教授或助教咨询。*
- *登录以下网站查找常见问题的回答：http://www.palisade.com。*
- *通过电子邮件或传真与我们的技术支持部联系。*

NeuralTools 系统要求

用于 Microsoft Excel for Windows 的 NeuralTools 5.7 系统要求包括：

- *配备硬盘的 Pentium 个人计算机或速度更快的个人计算机。*
- *Microsoft Windows 2000 SP4、Windows XP 或更高版本。*
- *Microsoft Excel 2000 或更高版本。*

安装说明

一般安装说明

安装程序会将 NeuralTools 系统文件复制到您指定的硬盘目录中去。

要在 Windows 2000 或更高版本的系统上运行安装程序：

1) *将 NeuralTools 或 DecisionTools Suite CD-ROM 插入您的 CD-ROM 驱动器中*
2) *依次单击“开始”按钮、“设置”和“控制面板”*
3) *双击“添加/删除程序”图标*
4) *在“安装/卸载”选项卡上，单击“安装”按钮*
5) *按照屏幕上显示的“安装”说明进行安装*

如果您在安装 NeuralTools 时遇到问题，请确认尝试安装 NeuralTools 的驱动器是否有足够的空间。释放足够的空间后，尝试重新运行安装程序。

从计算机上移除 NeuralTools

如果您希望从计算机上移除 NeuralTools，请使用控制面板的“添加/删除程序”实用程序，并选择 NeuralTools 条目。

设置 NeuralTools 图标或快捷方式

在 Windows 任务栏中创建快捷方式

在 Windows 中，安装程序会自动在任务栏的“Programs\Palisade DecisionTools”菜单中创建 NeuralTools 命令。但是，如果在安装过程中出现问题，或者希望以后再以手动方式设置，请遵循以下说明。请注意，以下所列说明适用于 Windows XP Professional。其他操作系统的说明可能不同。

1) *单击“开始”按钮，然后将光标指向“设置”。*
2) *单击“任务栏和「开始」菜单”，然后单击“「开始」菜单”选项卡。*
3) *依次单击“自定义”、“添加”和“浏览”。*
4) *找到NeuralTools.EXE 文件，单击该文件，然后再单击“确定”。*
5) *单击“下一步”，然后双击您希望显示程序的菜单。*
6) *键入名称“NeuralTools”，然后单击“完成”。*
7) *在所有打开的对话框中单击“确定”。*

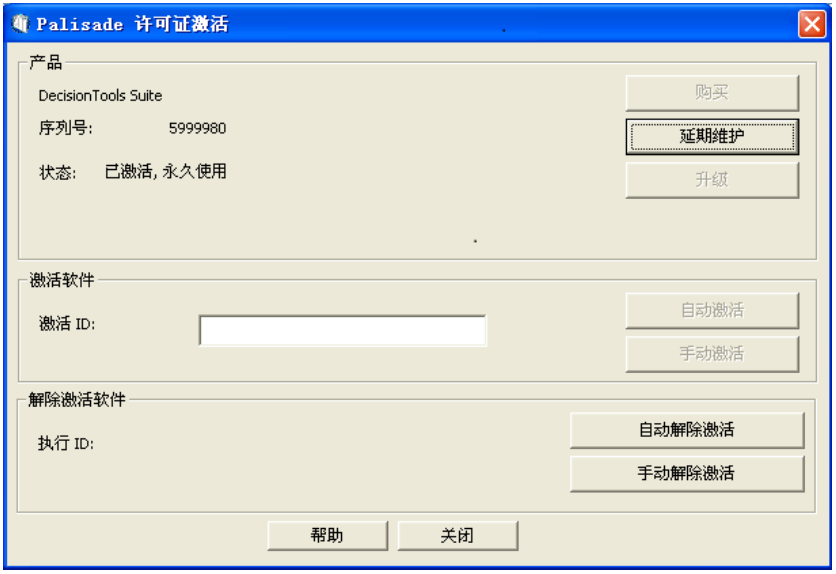
DecisionTools Suite

NeuralTools 可以与 DecisionTools Suite 配合使用，后者是 Palisade Corporation 提供的一套风险与决策分析产品。NeuralTools 的默认安装程序会将 NeuralTools 放入“Program Files\Palisade”主目录的一个子目录内。这与通常将 Excel 安装至“Microsoft Office”目录的一个子目录非常相似。

“Program Files\Palisade”目录的一个子目录将成为 NeuralTools 目录（默认名称为 NeuralTools5）。该目录包含 NeuralTools 插件程序文件 (NEURALTOOLS.XLA)、模型示例以及运行 NeuralTools 所需的其他文件。“Program Files\Palisade”的另一个子目录是“SYSTEM”目录，该目录包含 DecisionTools Suite 中的每个程序所需的文件，包括普通帮助文件和程序库。

软件激活

激活是一个一次性许可确认过程，是使您的 NeuralTools 软件作为完全获得许可的产品运行所必须执行的步骤。在您打印/通过电子邮件发送的发票中有一个**激活代码**，这是用破折号分隔的序列，例如“19a0-c7c1-15ef-1be0-4d7f-cd”。如果在安装过程中输入激活代码，则在首次运行软件时软件即被激活，而无需再采取其他操作。如果希望在安装后激活软件，请选择 NeuralTools“帮助”菜单中的“许可证激活”命令，然后在显示的 **Palisade 许可证激活**对话框中输入您的激活代码。



常见问题

1) 如果我的软件未激活，该如何处理？

如果在安装过程中没有输入激活代码，或者安装的是试用版，则软件将作为具有时间和/或使用次数限制的试用版运行，而必须使用激活代码激活才能作为完全获得许可的产品运行。

2) 在激活之前我可以使用该产品多长时间？

未激活的软件可以运行十五天。产品的所有功能均具备，但每次启动程序时，都会显示“许可证激活”对话框，提醒您激活软件，并显示剩余的试用时间。如果 15 天试用期终止，则软件要求激活后才能运行。

3) 我如何检查激活状态？

您可以通过 NeuralTools“帮助”菜单“许可证激活”命令来查看“许可证激活”对话框。激活的软件显示**已激活**状态，试用版软件则显示**未激活**状态。如果软件未激活，则会显示允许软件运行的剩余时间。

4) 我如何激活软件？

如果没有激活代码，您可以单击“许可证激活”对话框中的“购买”按钮来获得激活代码。通过在线购买可以立即获得激活代码，并且提供一个在需要重新安装时用于下载安装程序的可选链接。如果通过电话购买，请致电本章**与 Palisade 联系**一节中所列的当地 Palisade 办事处。

通过互联网或电子邮件激活软件：

如果具有互联网访问权限，则如何激活软件

在 Palisade“许证可激活”对话框中，键入或粘贴激活代码，然后按“自动激活”按钮。几秒钟后会显示一条成功消息，“许可证激活”对话框将反映软件的激活状态。

如果没有互联网访问权限，则如何激活软件

通过电子邮件自动激活需要执行以下几个步骤：

1. 单击**手动激活**，显示 request.xml 文件，您可以将该文件保存在磁盘上或复制到 Windows 剪贴板上。（建议您记下 request.xml 文件在计算机中的位置。）
2. **将此 XML 文件复制或附加**到电子邮件中，然后将其发送至 *activation@palisade.com*。您很快会在电子邮件的回复地址收到一封自动回复邮件。
3. **将回复邮件中的 response.xml 附件保存到**硬盘驱动器中。
4. 单击“Palisade 许可证激活”对话框中的**处理**按钮，并浏览至 response.xml 文件。选择该文件，并单击“确定”。

此时会显示一条成功消息，“许可证激活”对话框将反映软件的激活状态。

5) 我如何将软件许可转移到另一台计算机？

转让许可（或称为**转换主机**）可通过“Palisade 许可证激活”对话框进行，但需要执行以下两个步骤：在第一台计算机上*解除激活*，然后在第二台计算机上*激活*。一种典型的转换主机的情况是将 NeuralTools 从办公室个人计算机上转到便携式计算机上。如果需要将许可从*计算机 1* 转换至*计算机 2*，则确保两台计算机均已安装该软件，并且在解除激活/激活转换主机的过程中保持互联网连接。

1. 在*计算机 1* 上，单击“解除激活软件”对话框中的**自动解除激活**。等待显示成功消息。
2. 在*计算机 2* 上，单击“激活软件”对话框中的**自动激活**。等待显示成功消息。

如果计算机没有互联网访问权限，您可以按照上述类似说明通过自动化电子邮件流程来进行主机转换。

6) 我拥有互联网访问权限，但仍然无法自动激活/解除激活软件。

您的防火墙必须设置为允许 TCP 访问许可授权服务器。如果是单个用户（无网络安装），则为：
http://service.palisade.com:8888（TCP 端口 8888 位于 http://service.palisade.com）。

第二章： NeuralTools 概述

概述 15

为什么要使用神经网络？ 15
NeuralTools 和神经网络 15
NeuralTools 菜单和工具栏 16
数据集和数据集管理器 16
变量类型 17
多范围数据 17
训练神经网络 18
综合进行训练、检验和预测 18
网络配置 19
训练预览 19
训练过程 20
训练报表 21
检验网络 22
检验报表 22
预测 23
预测结果 23
实时预测 24
NeuralTools 报表和图表 25

概述

NeuralTools 可在您熟悉的环境 Microsoft Excel 里为您提供强大的神经网络功能。NeuralTools 的各种程序，如：定义数据集、训练和检验神经网络，以及使用已训练的网络预测值等，均可在您的 Excel 数据上运行，还可在 Excel 中创建分析报表和分析图。

为什么要使用神经网络？

神经网络能学习复杂的数据关系。它们通过模拟人脑的功能，来辨别数据模式，然后对向其提供的新数据进行推断预测。使用神经网络解决的问题可分为两大类

- **分类问题：**试图确定一个未知项目属于哪类范畴的问题。例如：医疗诊断和还贷能力预测。
- **数值问题：**需要预测具体数值结果的情况。例如：股价预测和未来某段时间的销售量预测。

NeuralTools 附带各种示例，以显示如何应用神经网络解决不同的预测问题。**NeuralTools\Examples** 文件夹中包含以 Excel 工作簿的形式提供的示例。

NeuralTools 和神经网络

可使用 NeuralTools 通过以下四步建立和使用神经网络：

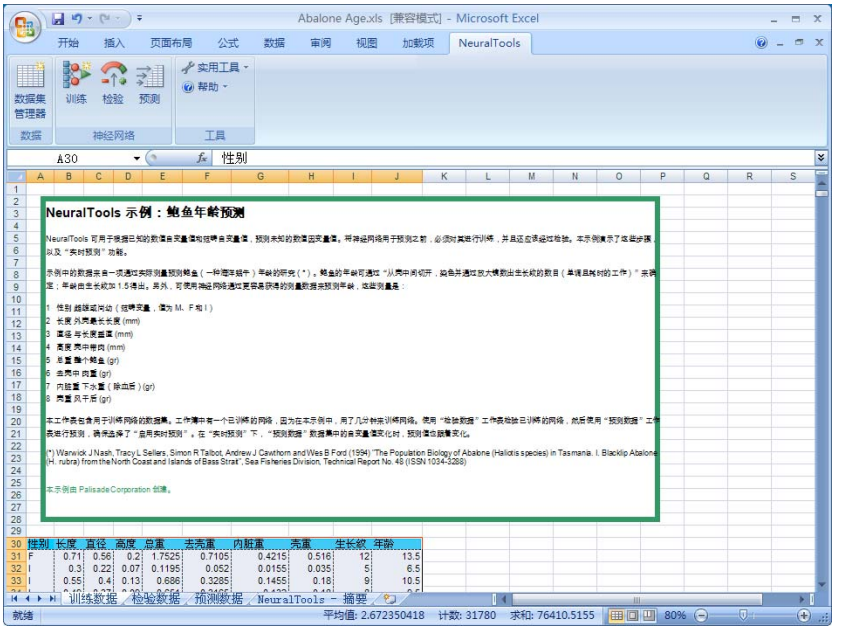
- **数据准备** – 您在 NeuralTools 中使用的数据在数据集中进行定义。使用**数据集管理器**设置数据集，以供您的神经网络重复使用。
- **训练** – 通过训练，即可从某个数据集中生成一个神经网络，该数据集须包括附带已知输出值的案例。这些数据通常由输出值/因变量值已知的历史案例组成。
- **检验** – 通过检验，可检测已训练的神经网络在预测已知输出值方面的表现。用于检验的数据通常是历史数据中的一个子集。该子集在训练网络时没有使用。检验后，用统计量来衡量网络的性能，如：正确预测已知答案的比例。
- **预测** – 使用一个已训练的神经网络预测未知的输出值。该网络经过训练和检验后，即可根据需要用于预测新案例数据的输出项。

训练和检验是一个不断重复的，有时也是很耗时的过程。通常您可采用不同的设置多训练几次，以生成检验效果最佳的神经网络。有了“最佳网络”后，您即可立即将其用于预测。

现在，我们来了解 NeuralTools 如何在 Excel 中工作，以及如何定义数据集并使用这些数据集训练和检验神经网络。然后，我们将使用已训练的网络来预测未知输出值。

NeuralTools 菜单和工具栏

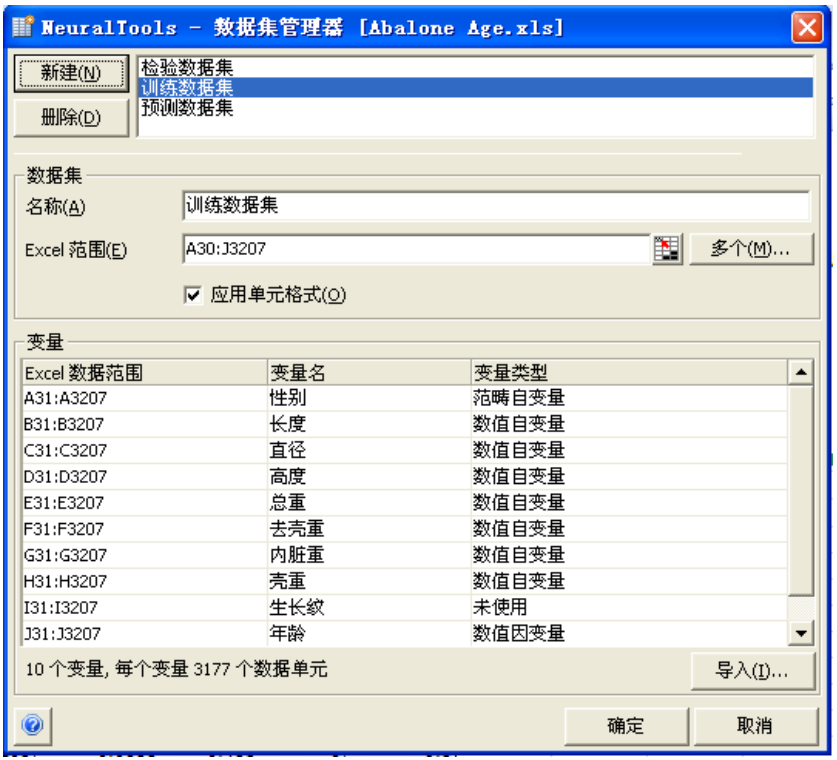
安装 NeuralTools 后，其菜单和命令就会嵌入 Excel 2003 或更早版本的 Excel 菜单栏。另外还会显示一个 NeuralTools 工具栏。该菜单显示：1) 定义数据集中数据的命令， 2) 训练和检验神经网络的命令，以及 3) 使用已训练的神经网络预测值的命令。在 Excel 2007 中，所有命令均通过 NeuralTools 功能导航栏提供。



数据集和数据集管理器

NeuralTools 中的数据围绕**案例**和**变量**构造。您使用的数据集或一组统计变量须按连续列排放，且数据集的第一行为变量名。数据集中的每一行都是一个**案例**。每个案例都有一组自变量值和一个已知的因变量输出值，或者没有因变量输出值。NeuralTools 的工作是预测案例的未知变量输出值。

NeuralTools **数据集管理器**可定义数据集、变量和案例。这样，您即可使用这些预定义变量来训练和检验神经网络，无需一次次重复选择您想要分析的数据。您可以把所有已知的历史案例放在一个数据集中，您想预测结果的案例放在另一个数据集中。您还可以将所有数据（已知历史数据和您想预测的数据）放在同一个数据集中。



数据集中的每个变量都有一个名称和一个与之关联的 Excel 单元范围。该范围中的每一列包含另一个变量的数据。一个数据集可包含多个单元区块，可将数据放在同一个工作簿的不同工作表里。

定义数据集时，NeuralTools 会尝试识别当前选中的 Excel 单元区块边缘的变量。这样可方便快捷地设置数据集，使其首行为变量名，变量按列排放。

变量类型

在 NeuralTools 中，变量可以是自变量或因变量，也可以是数值变量或范畴变量（如：*是*或*否*；*红*、*绿*或*蓝*）。数据集管理器会尝试识别您的数据集中每个变量的类型，但您可以用自己的选择进行覆盖。

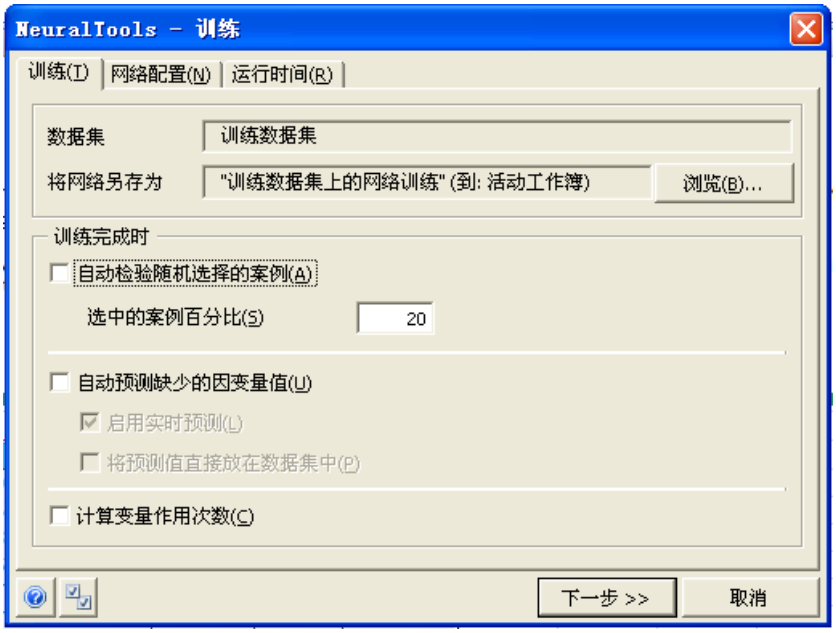
多范围数据

在 Excel 2003 或更早版本的工作表中，一列最多可存放某个变量的 65,536 个数据点。在变量所拥有的值超过这个数且您不采用 Excel 2007 的情况下，NeuralTools 允许将多个单元范围分配给一个数据集。换言

之，您可将一个数据集“复制”到多个工作表中，分配不同工作表中相同的列，以存放某个数据集的所有数据。

训练神经网络

定义了数据集且其中包含历史值已知的案例之后，您即可使用这些数据训练神经网络。有不同的选项来确定 NeuralTools 生成的网络类型。根据数据的性质，采用不同的网络选项会生成性能更好的已训练网络（即：在预测答案方面更出色的网络）。在训练后进行的检验过程可准确地衡量已训练的网络在预测输出值方面的表现。



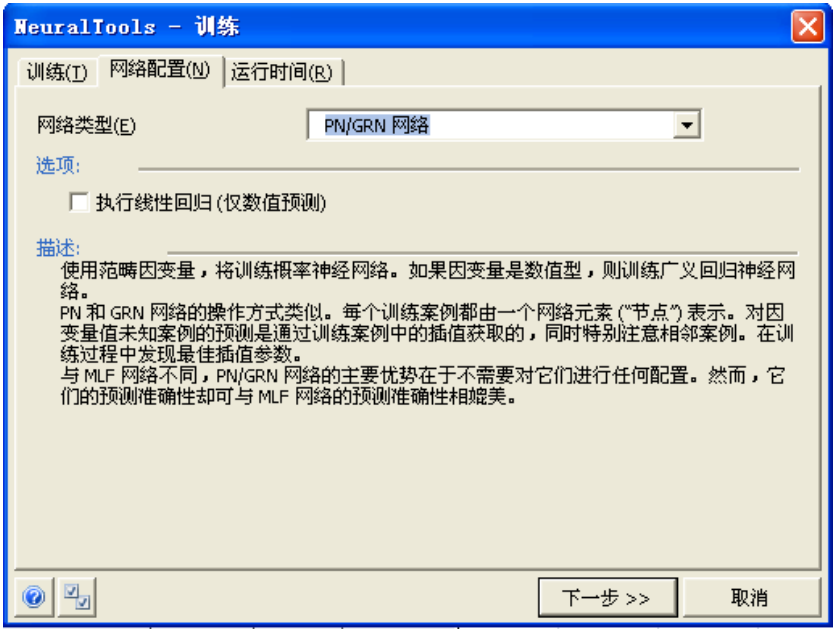
训练神经网络以及进行检验和预测时，要求您指定一个数据集，其中包含训练时要使用的数据。NeuralTools 会将已训练的网络直接保存在您的工作簿中，也可选择保存到硬盘上的一个文件中。

综合进行训练、检验和预测

如果您的所有数据在一个数据集中（包括已知历史数据和您不知道输出值的新数据），可使用 NeuralTools 训练和检验网络，然后预测输出值，所有操作一步完成。选择保留一定比例的历史数据进行检验（上一页显示为 20%），然后选择对缺少因变量值的案例自动预测输出值。这样，您只需一步操作即可快速得到您需要的答案。

网络配置

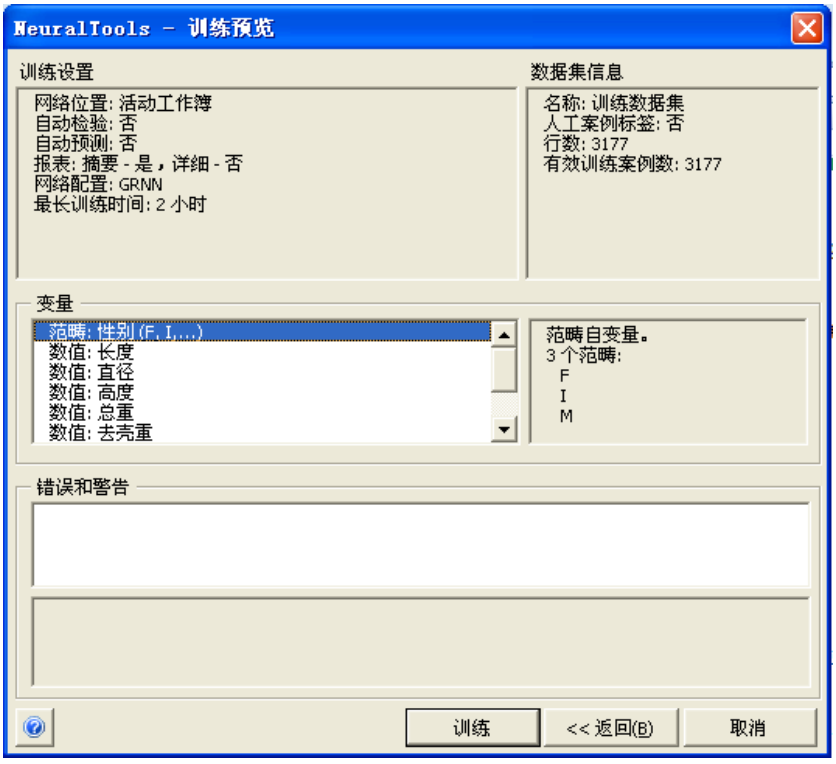
NeuralTools 支持不同的神经网络配置，以提供最佳预测。对于分类/范畴预测（其中，因变量是一个范畴类型），可使用两种网络：**概率神经网络 (PNN)** 和**多层前馈网络 (MLF)**。可使用 MLF 网络以及与 PNN 网络密切相关的**广义回归神经网络 (GRNN)** 进行数值预测。



NeuralTools 提供**最佳网络**搜索，使选择网络配置变得十分容易。选择该搜索后，NeuralTools 会训练和检验各种神经网络配置以生成一种可对您的数据进行最佳预测的配置。最佳配置取决于检验数据，因此，对于最佳网络搜索，需要在“训练”选项卡上选择“自动检验”选项。

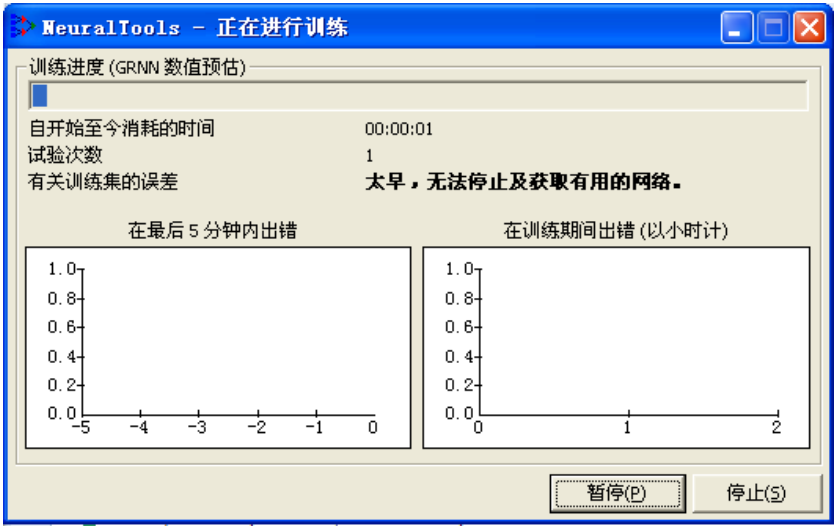
训练预览

选定训练和网络配置选项后，NeuralTools 会在训练网络期间预览要执行的操作。由于训练是最耗时的神经网络建模过程，因此在进行前查看训练设置非常有用。NeuralTools 会尝试在您的数据中找出问题，以便您在进行训练前加以更正。



训练过程

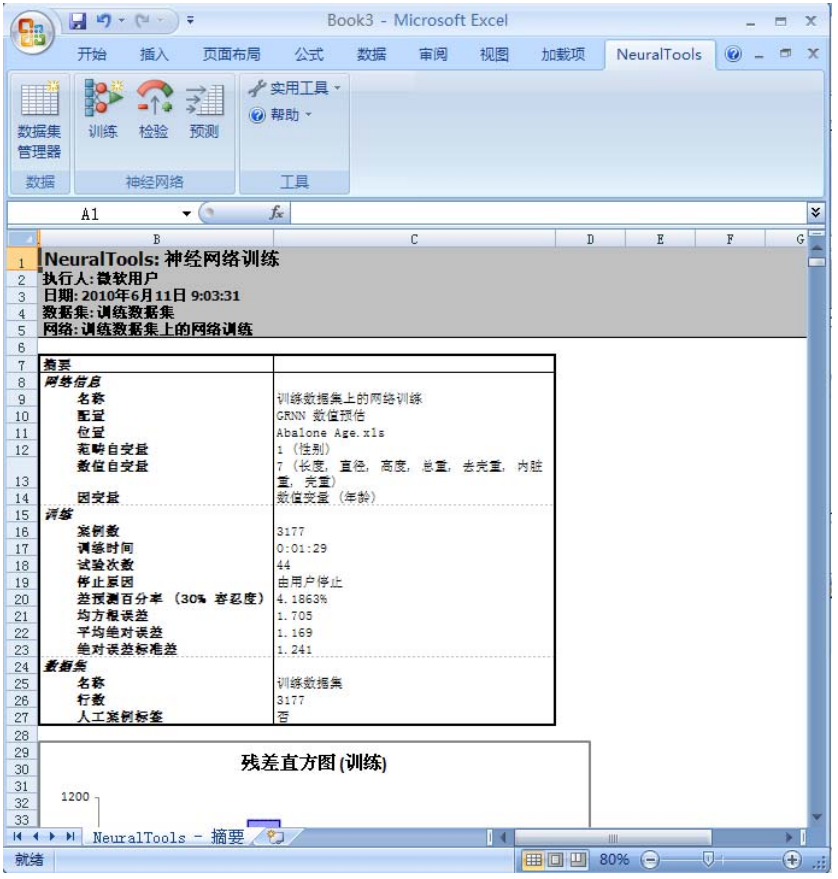
当 NeuralTools 用您的数据训练一个神经网络时，会报告其训练表现。通常，该网络会随着训练的深入而变得越来越好，因为 NeuralTools 生成的网络对数据的预测更准且错误更少。在训练期间，画面会不断更新，以显示 NeuralTools 的进展。



在达到您设置的任何一个停止条件（如：最长训练时间）时，训练会停止。如果您选择了自动检验网络或预测数据集中缺少的输出值，会在训练结束后执行。

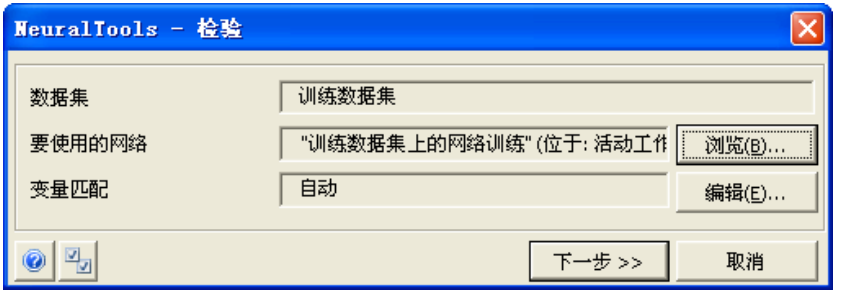
训练报表

训练报表显示已训练网络的表现。诸如**不良预测百分率**之类的统计量显示在这组训练中网络预测的输出值和已知实际值不相符的案例数。



检验网络

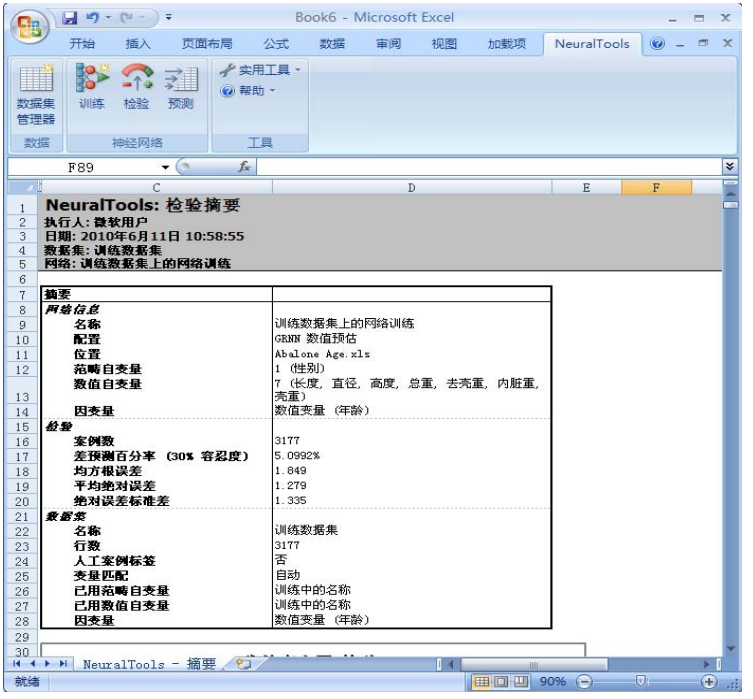
在检验过程中，会检验已训练的神经网络，以了解其在预测已知输出值方面的表现。检验数据通常是有已知输出值的历史数据中的一个子集。该子集在训练网络时没有使用。



如果检验数据在一个单独的数据集中，NeuralTools 会将检验数据集内的变量与训练数据中的变量相匹配。与训练相同，NeuralTools 会在运行前预览您的检验设置。

检验报表

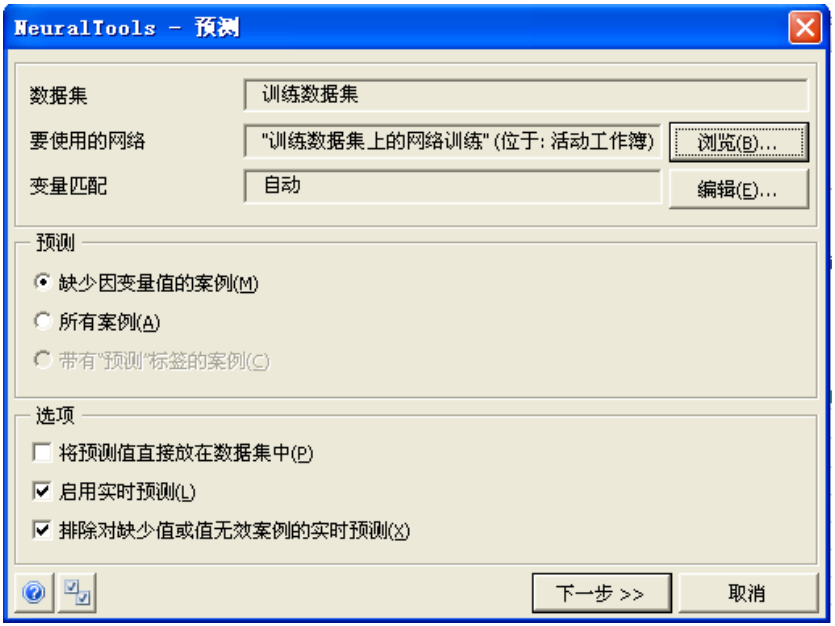
检验（以及预测）运行起来比训练快得多。NeuralTools 报告网络在预测检验数据中的已知答案方面的表现。这样有助于您了解该网络在用于输出值未知的案例时，是否是一种优良的预估工具。



预测

神经系统的最终用途是预测。您会将已训练的网络用于您不知道输出值、但想预测它的新案例。NeuralTools 提供两种预测方法：1) **预测值命令法**，适用于一个数据集内的案例；2) **实时预测法（仅企业版）**，在这种情况下，工作表中某个案例的自变量值可以输入，且 NeuralTools 将自动计算预测的输出值。

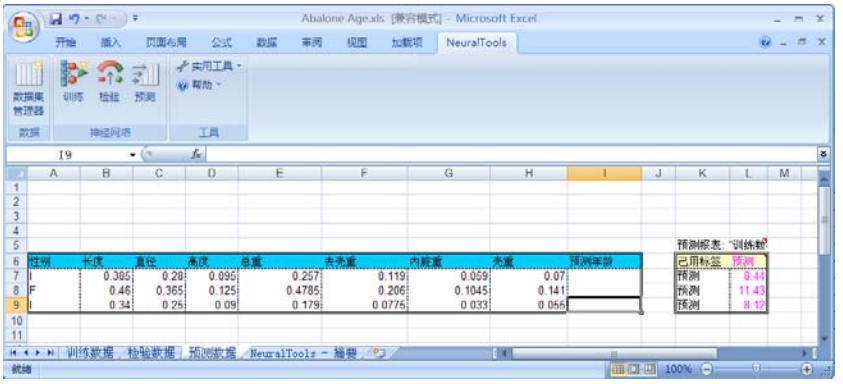
如果您想对某个数据集中的一组案例预测值，可使用预测对话框设置预测程序。您可仅预测缺少输出值的案例，并可选择启用“实时预测”以便修改您的数据，从而了解它对预测的影响。可使用经过不同训练的网络来了解预测值的变化。



同训练和检验一样，NeuralTools 首先预览将用于预测的数据和设置。然后，会在您的 Excel 工作表中报告预测结果。

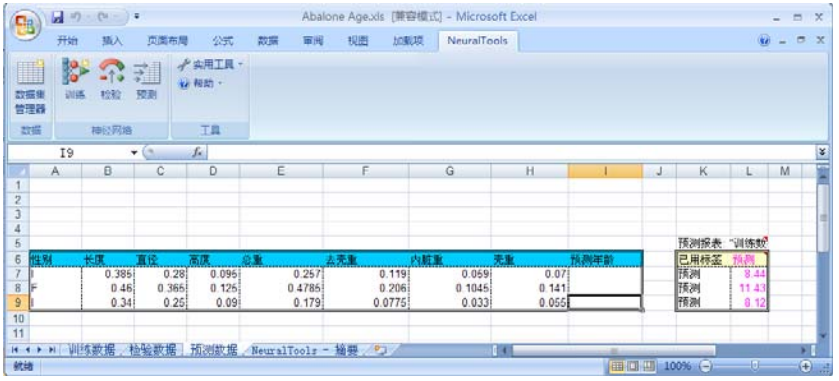
预测结果

预测的输出值在被预测案例的旁边显示，在以下画面中，预测值为紫色。



实时预测

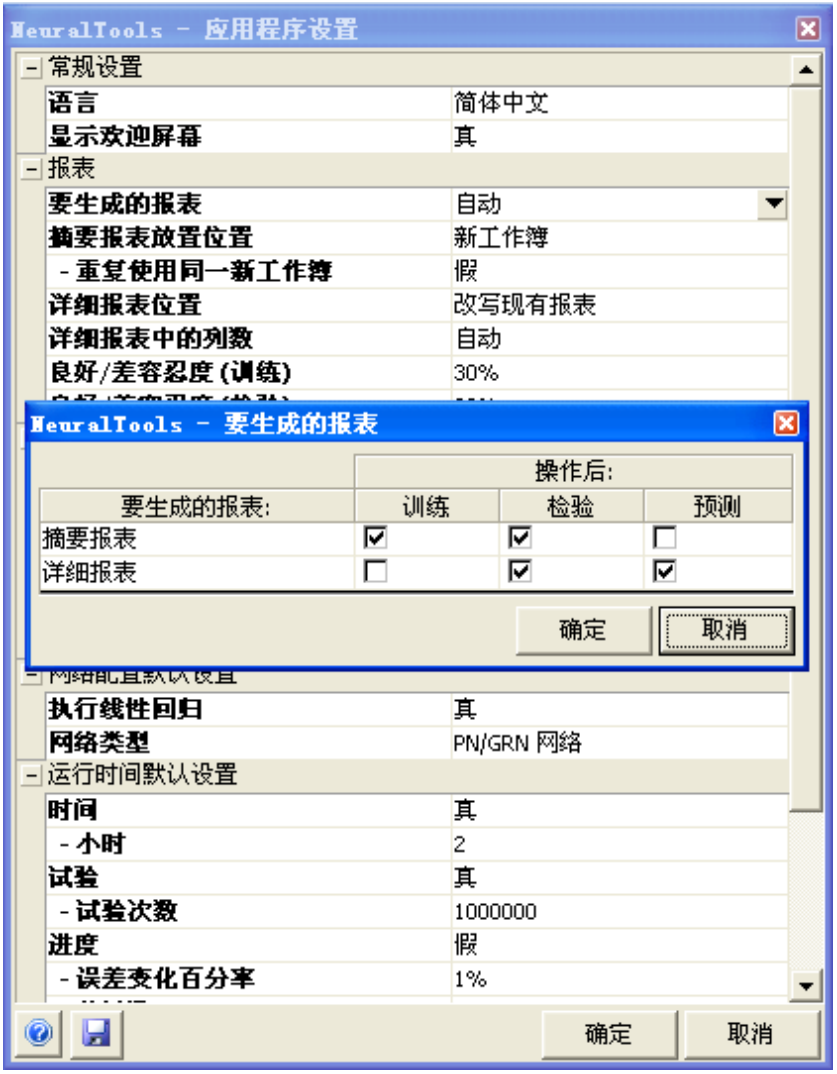
启用实时预测时，NeuralTools 会将一个 Excel 公式自动添加到显示预测值的单元中。这个公式生成预测值。因此，如果修改了案例的自变量值，则会自动重新计算预测值。使用实时预测时，您只需直接在 Excel 里为新案例输入数据，即可自动生成新预测而无须使用预测对话框。例如：如果以上工作表中第 7 行案例的自变量值更改如下，预测值则会自动更新。与所有工作表单元一样，您可以在任何 Excel 公式里引用 LivePrediction 单元。



（注：实时预测仅对企业版提供。）

NeuralTools 报表和图表

NeuralTools 可对训练、检验和预测创建摘要报表和详细报表。**摘要报表**显示在一个单独的工作表中，包含检验或训练的总体信息。**详细报表**则视不同的案例提供不同的信息，显示在报表所报告数据的旁边。此外，摘要报表的大多数信息均可在详细报表中找到，它们以注释的形式添加到标题单元；这种摘要报表称为**快速摘要**。



每当 NeuralTools 创建一张或多张图表时，总是将其和报表放在一起。图表以 Excel 格式创建，可使用标准 Excel 图表命令进行自定义。

NeuralTools 实用工具

提供两种实用工具来帮您管理在 NeuralTools 中为神经网络建模。一个是**神经网络管理器**，可使用它在工作簿和文件之间复制或移动已训练的神经网络。另一个是**缺失数据**实用工具，它有助于识别和更正您的数据集内缺失数据的案例。

将 NeuralTools 与 StatTools、Solver 和 Evolver 配合使用

NeuralTools 设计用于配合 **StatTools** 使用，后者是 Palisade 提供的一款 Excel 统计插件。两个产品共用一个数据集管理器；在 NeuralTools 中定义的数据集可使用 StatTools 进行分析，反之亦然。您可使用 StatTools 对 NeuralTools 中定义的数据集变量进行统计计算，还可对 NeuralTools 生成的预测值进行统计计算。

在 NeuralTools 中生成的详细报表可直接在 StatTools 中进行分析；这些报表会自动显示在 StatTools 数据集管理器的数据集列表中。这样便于使用 StatTools 获得 NeuralTools 摘要报表中没有的统计结果。例如：某个检验摘要报表中包含一个残差（定义为实际值与预测值之差）直方图。根据该直方图，残差大致呈正态分布。为了检验正态分布这个假设，可对详细报表中的残差变量采用 StatTools 的一种正态性检验。相关示例请见文件“Abalone Age Prediction with StatTools Analysis.xls”。

NeuralTools 的 Live Prediction（实时预测）功能便于了解自变量值的改变对预测的影响。凭借 Live Prediction，即可使用 Excel 中提供的其他工具来研究自变量和因变量之间的关系。

Solver – Excel 的内置优化器，可与 NeuralTools 的实时预测功能配合使用，从而对在 NeuralTools 中作出的预测计算最佳判定值。相关示例请见文件“Auto Loans with Solver.xls”。在该示例中，使用一个神经网络预测借款人是否会及时还款。可是，网络给出的置信度只有 60%。那么，可以使用 Excel 的 Solver 来确定一个放贷额度，使网络对此人的及时还款把握达到 90%。在此例中，该优化器会尝试不同的放贷额度，而 NeuralTools 则自动更新概率值。可使用基于 Palisade 遗传算法的优化器 **Evolver** 替代 Solver 找出答案。与 Solver 不同，Evolver 可处理有多个局部最优值的优化问题。

第三章：NeuralTools 参考指南

简介 29

参考：NeuralTools 图标 31

NeuralTools 工具栏 31

参考：NeuralTools 菜单命令 33

简介 33

对话框中的图标 34

命令参考 35

数据集管理器命令 35
训练命令 40
检验命令 51
预测命令 58

实用工具 61

应用程序设置命令 61
神经网络管理器命令 64
缺失数据实用工具命令 66

神经网络的更多信息 69

简介

『NeuralTools 参考指南』一章讲述 NeuralTools 使用的图标、命令和统计量函数。本章分为两部分：

1) *参考：NeuralTools 图标*
2) *参考：NeuralTools 菜单命令*

参考：NeuralTools 图标

NeuralTools 工具栏

NeuralTool 图标用于定义带有案例和变量的数据集，然后用这些数据创建和使用神经网络。在 Excel 2003 及更早版本上，NeuralTools 图标出现在 Excel 工具栏上（即：Excel 上的一个自定义工具栏）；在 Excel 2007 上，该图标则出现在一个功能导航栏上。这一部分简要介绍了每个图标，大致说明了它们执行的功能，以及与之关联的等效菜单命令。在 Excel 2007 中，所有命令均通过 NeuralTools 功能导航栏提供。

以下图标显示在 Excel 2003 及更早版本的 NeuralTools 工具栏上和/或 NeuralTools 对话框中。

图标	***执行的功能和等效命令***
	定义数据集和变量，或者编辑或删除现有的数据集和变量 *等效命令：数据集管理器命令*
	训练神经网络 *等效命令：训练命令*
	检验神经网络 *等效命令：检验命令*
	使用已训练的网络预测值 *等效命令：预测命令*
	运行神经网络实用工具 *等效命令：实用工具命令*
	显示 NeuralTools 帮助文件 *等效命令：帮助命令*

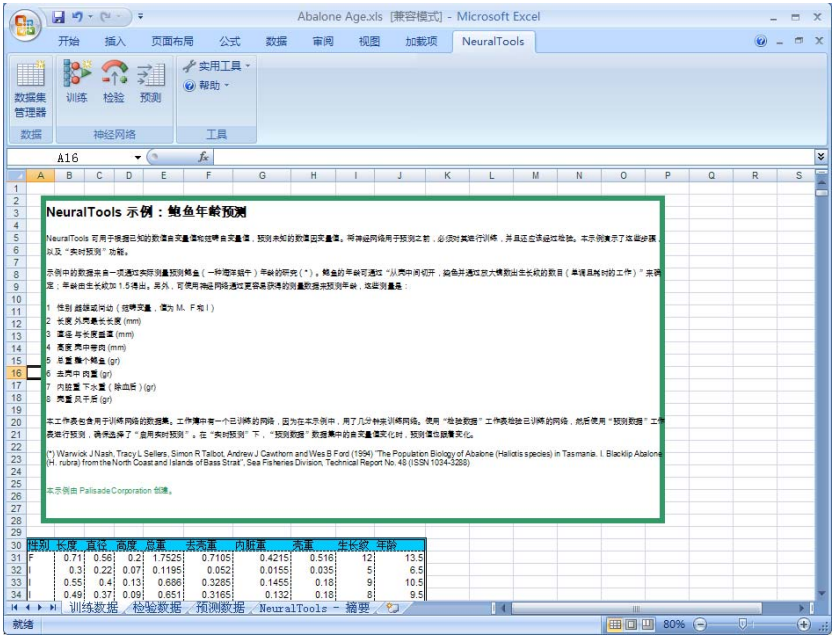
以下图标显示在 Excel 2007 的 NeuralTools 功能导航栏。

图标	***执行的功能和等效命令***
	定义数据集和变量，或者编辑或删除现有的数据集和变量 *等效命令：数据集管理器命令*
	训练神经网络 *等效命令：训练命令*
	检验神经网络 *等效命令：检验命令*
	使用已训练的网络预测值 *等效命令：预测命令*
	运行神经网络实用工具 *等效命令：实用工具命令*
	显示 NeuralTools 帮助文件 *等效命令：帮助命令*

参考：NeuralTools 菜单命令

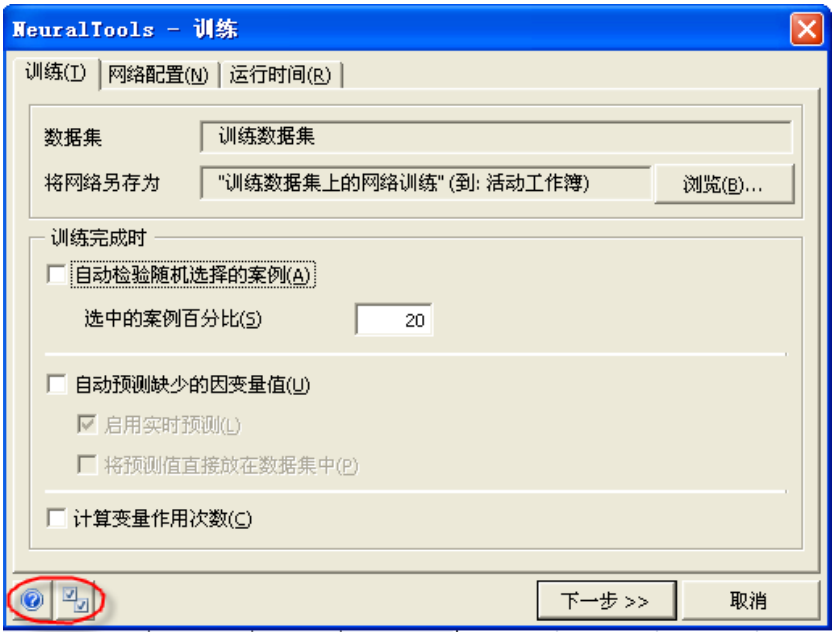
简介

『参考指南』的这一部分详细介绍了 Excel 中的 NeuralTools 菜单上或功能导航栏中显示的可用 NeuralTools 命令。将对出现在菜单上的命令进行介绍，从数据集管理器命令开始，逐一向后介绍。可使用 NeuralTools 图标执行许多可用的命令。本章的**参考：NeuralTools 图标**部分提供每个 **NeuralTools** 图标的等效命令。



对话框中的图标

在一个 NeuralTools 对话框中最多可以出现两个图标：**帮助图标**和**应用程序设置图标**。可使用帮助图标快速访问相关对话框的帮助主题。“应用程序设置”图标显示“应用程序设置”对话框，供您输入或编辑 NeuralTools 报表的设置，以及选择训练、预测和运行时间的默认设置。



命令参考

数据集管理器命令

定义 NeuralTools 数据集和变量，或者编辑或删除现有的数据集和变量

数据集管理器命令可定义数据集的案例和变量。定义了数据集后，即可将其用于神经网络的训练、检验和预测。可使用数据集管理器对话框添加或删除数据集、命名数据集，指定变量在数据集中的排放，以及在数据集中命名变量。

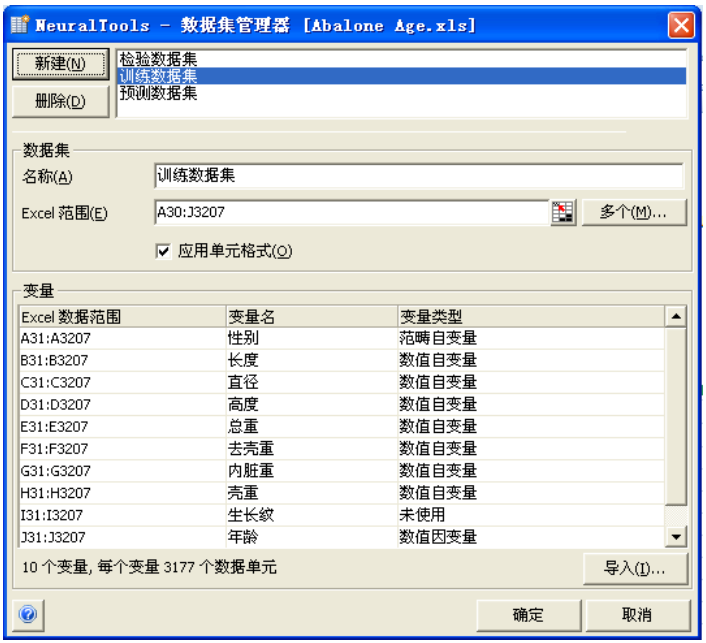
数据集和变量是什么？

NeuralTools 是围绕变量和案例来构建的。您使用的数据集或一组统计变量在 Excel 工作表中按连续列排放，且数据集的第一行为变量名。数据集中的每一行都是一个**案例**。每个案例都有一组自变量值和一个已知的因变量输出值，或者没有因变量输出值。

数据集中的每个变量都有一个名称和一个与之关联的 Excel 单元范围。一个数据集可包含多个单元区块，可将数据放在同一个工作簿的不同工作表里。

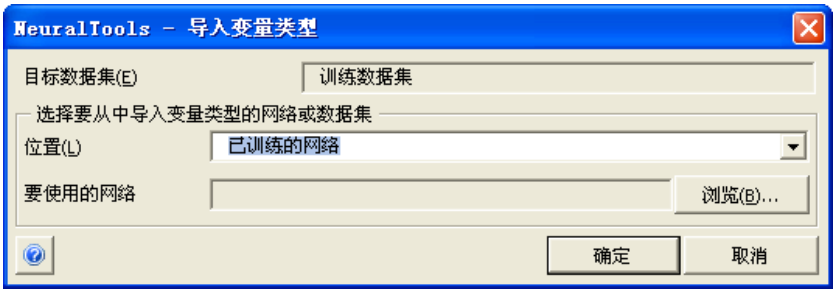
定义数据集时，NeuralTools 会尝试识别当前选中的 Excel 单元区块边缘的变量。这样可方便快捷地设置数据集，使其首行为变量名，变量按列排放。

数据集管理器对话框



数据集管理器对话框中的**数据集**选项包括：

- **新建/删除** – 添加新数据集或删除现有数据集。
- **名称** – 指定数据集的名称。
- **Excel 范围** – 指定与数据集关联的 Excel 范围。如果将多个单元范围分配给一个数据集，此条目会前缀有**多个**字样。
- **应用单元格式** – 添加可识别您的数据集的网格或颜色。
- **多个** – 单击数据集管理器对话框中的**多个**按钮，显示**多范围选择器**对话框。可在该对话框中输入组成多个单元范围数据集的各个单元范围。
- **导入** – 可从另一个数据集或已训练的神经网络将变量类型复制到此数据集。可使用“导入变量类型”对话框选择用于变量定义的位置和网络。



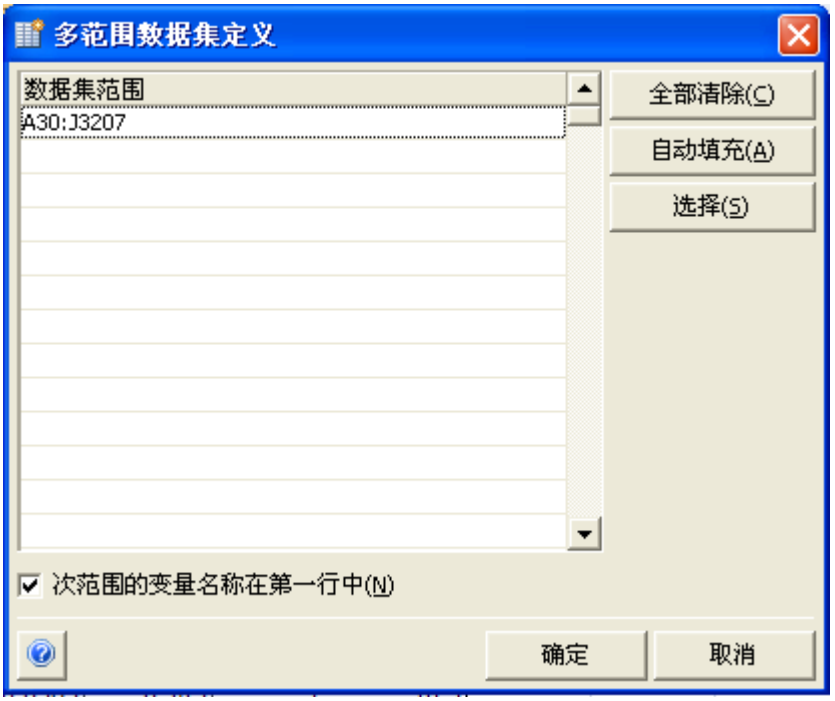
多范围数据集

NeuralTools 允许将不同工作表中的多个单元范围分配给一个数据集。可在以下情况下使用多范围数据集：

1) 在 Excel 2003 或更早版本的工作表中，如果一个数据集中的每个变量有 65,536 以上个数据点，则需要数据集在同一个工作簿的多个工作表上放置数据。

2) 变量的数据分布在一个工作簿的若干工作表上的多个区块里。

注：多范围数据集不能在一个工作表中定义，可在同一个工作簿的多个工作表上进行定义。

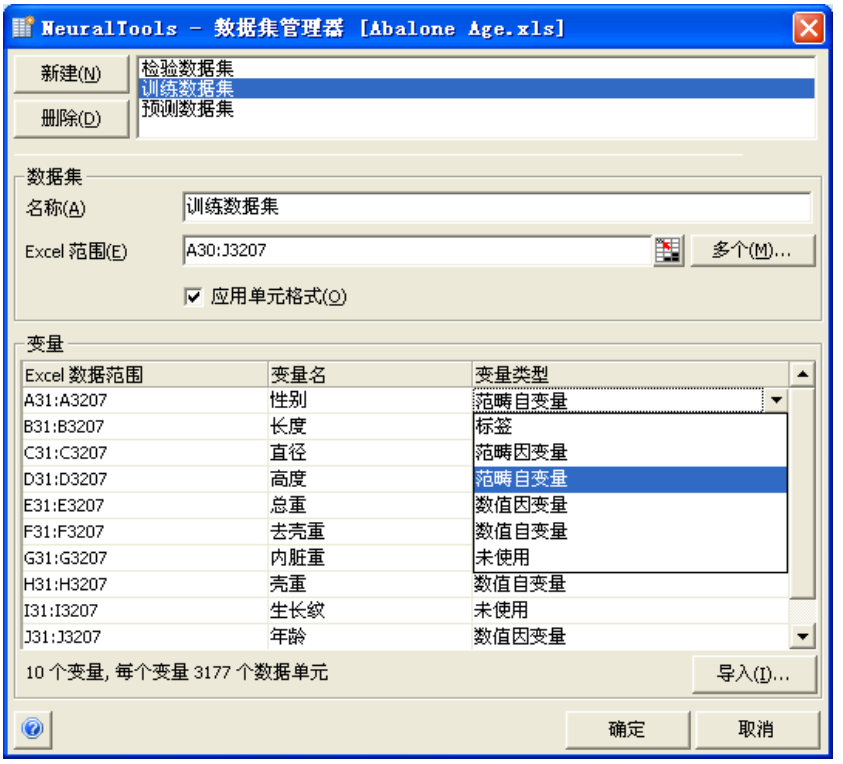


多范围选择器对话框中的选项包括：

- **全部清除** – 清除所有输入的范围。
- **自动填充** – 在当前工作簿中，对现有的所有工作表应用输入的第一个范围（第一行），并在网格中输入这些 *SheetName!CellRange* 引用。
- **选择** – 显示选择器，以选中作为数据集范围的一个单元区块。
- **次范围的变量名称在第一列（行）中** – 对于多范围数据集，对话框中会列出每一范围内描述每一列的变量名，或者仅列出描述第一个所选范围内各个列的变量名。第一个所选范围就是在多范围选择器对话框的第一行中输入的范围。

变量选项

数据集管理器对话框的网格中的每一行列出数据集中的变量，包括放置变量数据点的 **Excel 数据范围**、**变量名称**以及**变量类型**。



变量类型选项包括：

- **范畴因变量** – 可能值来自一组可能范畴的因变量或输出变量；例如：*是*或*否*，*红*、*绿*或*蓝*。
- **数值因变量** – 可能值为数值的因变量或输出变量。
- **范畴自变量** – 可能值来自一组可能范畴的自变量；例如：*是*或*否*，*红*、*绿*或*蓝*。
- **数值自变量** – 可能值为数值的自变量。
- **标签** – 可能值为“训练”、“检验”或“预测”的变量。此类变量用于识别数据集内用以进行训练、检验和预测的案例。
- **未使用** – 数据集中不会用于神经网络的变量。

标签变量的更多信息

标签变量是 NeuralTools 数据集中的一种特殊变量类型，用于识别数据集中用以进行训练、检验和预测的案例。如果您想将（用于进行网络训练、检验和预测的）所有数据都纳入一个数据集，使用标签变量则特别有用。有了标签变量，NeuralTools 即可根据标签变量的值选择用于训练、检验或预测的案例。通过更改标签变量值，您可使用其他案例再次训练网络，以了解网络性能如何变化。您还可以向数据集添加因变量值未知的新案例，并使用“预测”标签对其进行预测。标签变量仅可有三个可能值：

- **训练** – 指定该案例用于进行训练
- **检验** – 指定该案例用于进行检验
- **预测** – 指定该案例用于进行预测

注：如果您的数据集中有标签变量，“训练”对话框中的选项会改变。请参阅训练命令以了解更多信息。

数据集和变量的容量

在一个会话里，NeuralTools 允许：

- 一个工作簿中最多有 256 个数据集。
- 在 Excel 2007 中，每个数据集最多有 16384 个变量（较早版本的 Excel 中为 256 个变量）。一个数据集的所有数据必须在同一个工作簿中。
- 每个变量的数据点数量以及每个数据集中的案例数量仅受 Excel 2007 提供的存储量限制（较早版本的 Excel 为 16,777,216 个数据点）。

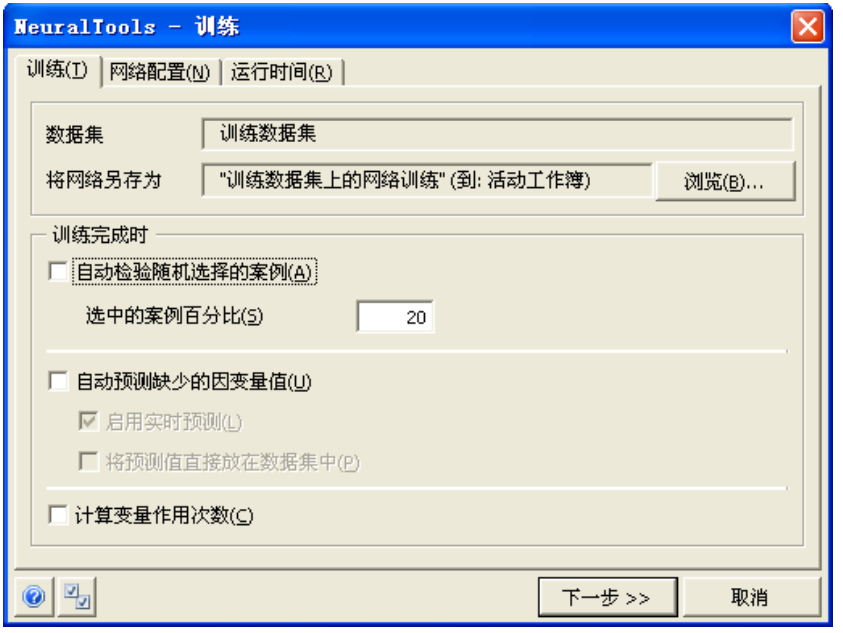
根据系统配置和使用的 Excel 版本，实际数据容量可能比上述少。Excel 本身的存储量限制也会影响数据的容量。

注：数据集管理器对话框列出当前工作簿（即：列在数据集管理器对话框标题的工作簿）中的所有数据集和变量。如需列出其他工作簿中的数据集，可激活 Excel 中所需的工作簿并显示数据集管理器对话框。

训练命令

指定用以训练神经网络的设置并进行训练

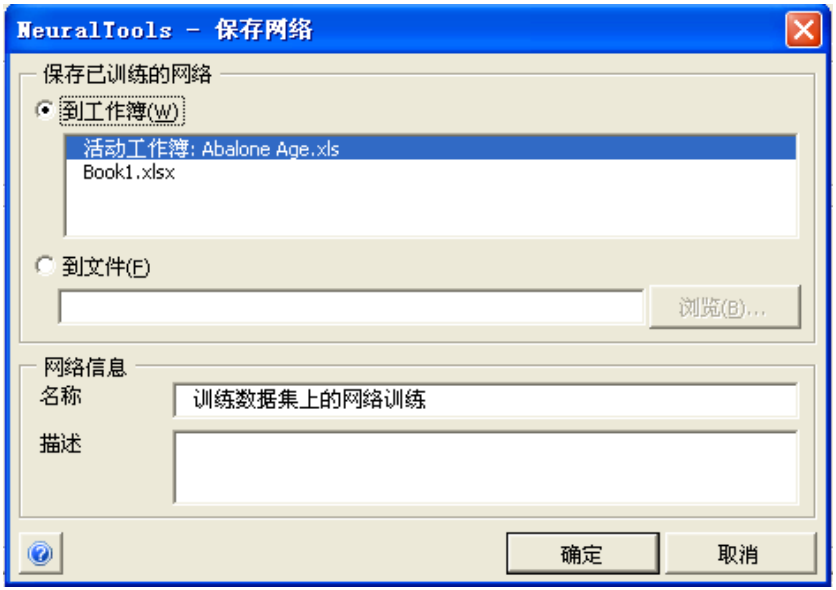
训练命令可 1) 在 NeuralTools 中指定用于训练某个神经网络的设置，以及 2) 开始训练一个网络。



训练选项卡

训练对话框中的**训练选项卡**指定用于训练一个神经网络的常用选项，它包括以下选项：

- **数据集** – 显示训练神经网络时使用的数据集。此数据集需要使用数据集管理器进行定义并位于当前工作表中。
- **将网络另存为** – 指定要训练的神经网络的名称和位置。可将神经网络保存到一个 Excel 工作簿中，也可保存到硬盘上的一个文件中。单击**浏览**... 以更改显示的名称和位置。



您还可以为要保存的神经网络输入名称和描述。

可使用**训练完成时**选项在训练后使用已训练的网络自动进行检验和预测。如果要检验和预测的数据与训练数据在一个数据集中，即可完成以上操作。

- **自动检验** – 指定：
  1) 在数据集中自动从训练中“排除”，以用于检验的案例 %比例
  2) 标签变量为“检验”的案例将用于检验。标签变量是在数据集管理器指定的一种变量类型。
- **自动预测缺失的因变量值** – 指定已训练的网络将用于预测以下一种情况下的因变量值：
  1) 缺失因变量值的案例，
     或者
  2) 标签变量为“预测”的案例。标签变量是在数据集管理器指定的一种变量类型。
- **启用实时预测** – 指定 NeuralTools 将公式放在显示因变量预测值的单元中，以计算预测值。有关实时预测的更多信息，请参阅本章的预测命令。
- **计算变量作用次数** – 指定 NeuralTools 计算训练数据集内的每个自变量在确定网络计算的预测值方面的相对作用。

变量作用分析是什么？

变量作用分析旨在衡量网络预测对自变量改变的敏感性。仅对训练数据进行这种分析。经过分析，每个自变量都会被赋予一个“相对变量作用次数”值，这些值是百分比值，加起来为 100%。对于给定变量来说，这个百分比值越低，该变量对预测的影响就越小。分析结果有助于选择一组新自变量，即能够更准确地进行预测的变量。例如，可去掉一些作用小的值并加入一些新变量。但是，须切记，作用分析的结果是相对于给定网络的。一个网络“学会”忽略某个给定变量，可能会使另一个网络也“学会”忽略它；但是，使用另一种网络进行的另一个训练可能会“发现”这个变量会对准确预测有重大贡献。在案例数很少的数据集和/或变量数很多的数据集中，变量对已训练网络所产生的相对作用会有很大的差异。另外，切记这些值是“相对的”。假设有两个自变量，一个赋值为 99%，另一个赋值为 1%。这表示后者远不如前者重要，但并不意味着后者不重要，需要高度准确的预测时尤其如此。

关于变量作用分析方面，其他需要注意的地方包括：

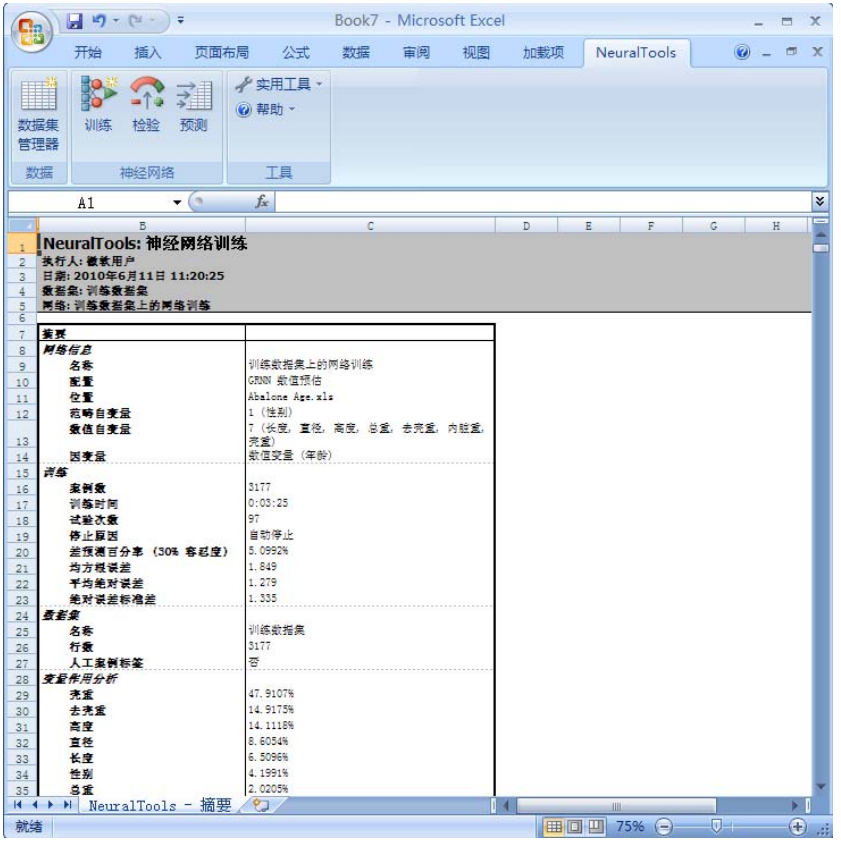
1) 分析中仅包括训练数据。（如果使用自动检验或自动预测，则相关案例不在分析之列。原因是它们可能会有训练范围之外的数值，这会使分析结果更难预测。）

2) 对于给定的范畴自变量，会针对每个案例分析该变量的所有有效范畴，并衡量对预测值带来的变化。（对于范畴预测，没有预测数值，但有一个范畴预测所依据的原始网络输出数值，分析会使用这些输出数值。）

3) 对于给定的数值自变量，会针对每个案例分析该变量上下限范围内的所有训练值，以衡量对预测值带来的变化（或者，对于范畴预测，则是衡量对原始输出数值带来的变化）。

变量作用分析的目的不是支持定论，如：确凿表明给定变量不相关。而是帮助寻找一组最佳的自变量：分析的结果可能告诉我们，某个给定变量似乎极不相关，所以，值得尝试不用此变量对网络进行训练。

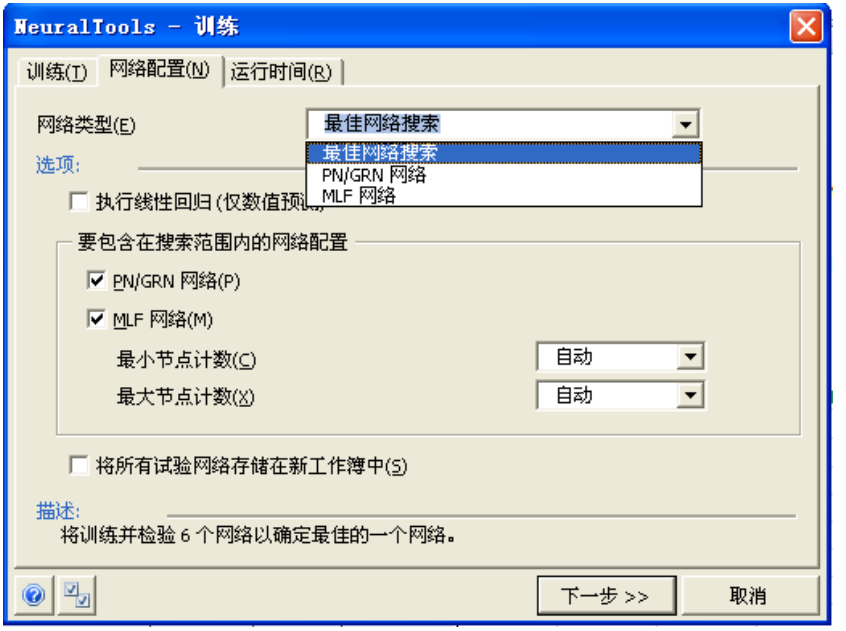
变量作用分析的结果显示在训练摘要报表中：



网络配置选项卡

可使用训练对话框中的**网络配置选项卡**来选择将用您的数据进行训练的神经网络类型。您可以选择一种特定的网络配置或选择**最佳网络**搜索，对于后者，NeuralTools 会检验可能采用的各种配置以确定对您最佳的配置。

NeuralTools 支持不同的神经网络配置，以提供最佳预测。对于分类/范畴预测，可使用两种网络：**概率神经网络 (PNN)** 和**多层前馈网络 (MLF)**。可使用 MLF 网络以及与 PNN 网络密切相的**广义回归神经网络 (GRNN)** 进行数值预测。有关可用网络配置更多技术信息，请参阅**神经网络的更多信息**一节。



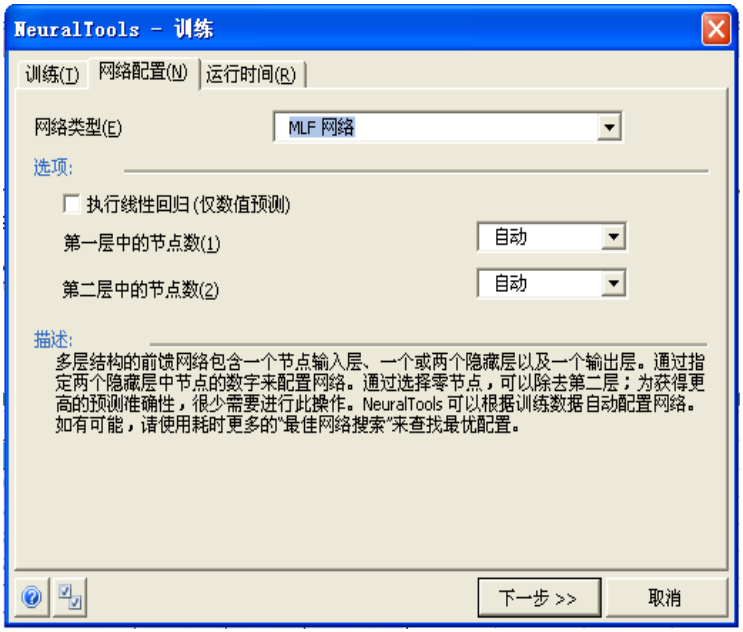
网络配置选项卡包括以下内容：

- **网络类型** – 选择训练时使用的网络类型，或者选择“最佳网络“搜索。网络配置选项卡的**选项**视选择的网络类型而有所不同。可用的网络类型有：

    1) **最佳网络搜索**。在“最佳网络搜索”中，NeuralTools 检验所有选择的网络配置，包括节点数在输入的上下限范围内的 PNN/GRNN 以及 MLFN 网络。数据的最佳配置根据检验数据的误差来确定。如果选择了**将所有试验网络存储在新工作簿中**，您即可从一个工作簿中单独载入每个已检验的网络（无论其是否为最佳网络），并在训练后使用它进行预测；还可为每个网络提供一个完整的检验摘要报表。

2) **PNN/GRNN 网络。**这两类网络无需为训练选择其他选项；因此，在安装 NeuralTools 时，这一设置为默认值。如果您的数据有数值输出值，会训练 GRNN 网络；而如果您的数据有范畴输出值，则会训练 PNN 网络。

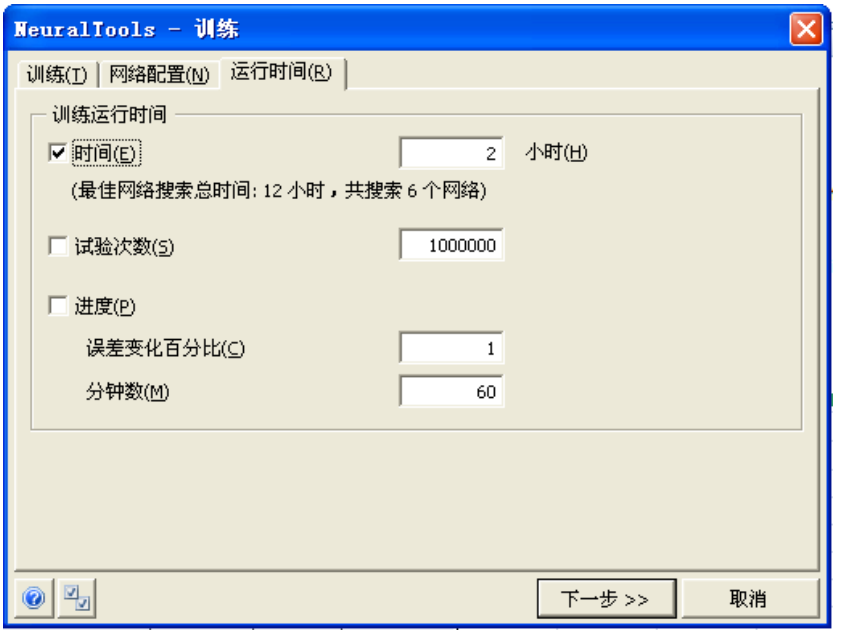
3) **MLF 网络。**一种有一个或两个隐含节点层的多层前馈网络 (MLFN)。



通过为第二层选择零节点，即可除去该层。找出 MLFN 网络最佳配置的最可靠方法是使用“最佳网络搜索”选项，而不是训练一个 MLFN 网络的选项。如果没有足够的时间进行“最佳网络搜索”，建议将“节点数”的值保留为“自动”。

运行时间选项卡

可使用训练对话框中的**运行时间选项卡**输入训练的停止条件。如果没有选择停止条件，训练最终还是会停止；对于 PNN/GRNN 网络，这一时间相对较短，而对于 MLF 网络则较长。可以采用的一种方法是选择不停止条件，在没有更多时间进行训练的时候，单击”训练进度“对话框中的”停止“按钮。使用“最佳网络搜索”时，必须定义训练一个网络的时限，以确保搜索算法不会在某种配置上花费太多的时间。可综合使用所提供的这三种停止条件，指定在达到任意一个条件时 NeuralTools 将停止。



“训练运行时间”选项包括：

- **时间** – 为训练一个网络指定一个固定的时限。只要算法确定不可能再有任何进展，即便没有到指定时间，训练也会停止。如果使用最佳网络搜索，每个被检验的网络配置均按输入的时间进行训练。
- **试验次数** – 指定 NeuralTools 在停止之前执行试验的次数不得超过指定次数。对于多层前馈网络，一次“试验”是神经元之间连接的一个“权重”赋值；训练是对可生成最佳预测的权重所进行的智能搜索。对于概率神经网络和广义回归神经网络，一次试验是变量的一个“平滑因子”赋值。训练是搜索最佳平滑因子。
- **进度** – 指定如果 NeuralTools 不能在指定的时间内将误差统计量至少减少到输入的百分比，则 NeuralTools 将停止。

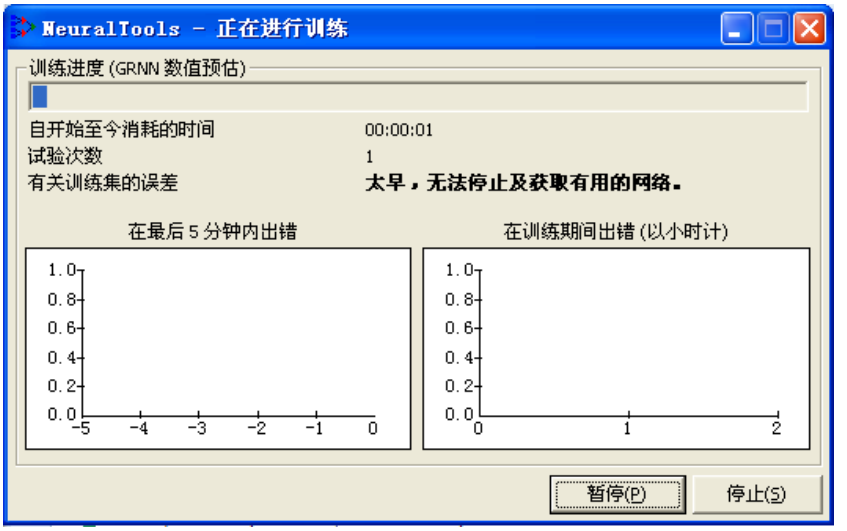
训练预览对话框

训练预览对话框在开始训练前显示当前网络训练设置，以及在数据中发现的错误。通过查看这一对话框的内容，您可根据 NeuralTools 的报告了解所有选定的训练假设。**错误和警告**部分描述了 NeuralTools 在数据和设置中发现的问题，如有必要，您可在花时间进行训练之前加以更正。

训练进度窗口

训练进度窗口在进行网络训练时报告训练状态。图表详细说明了 NeuralTools 如何改善网络和减少误报。

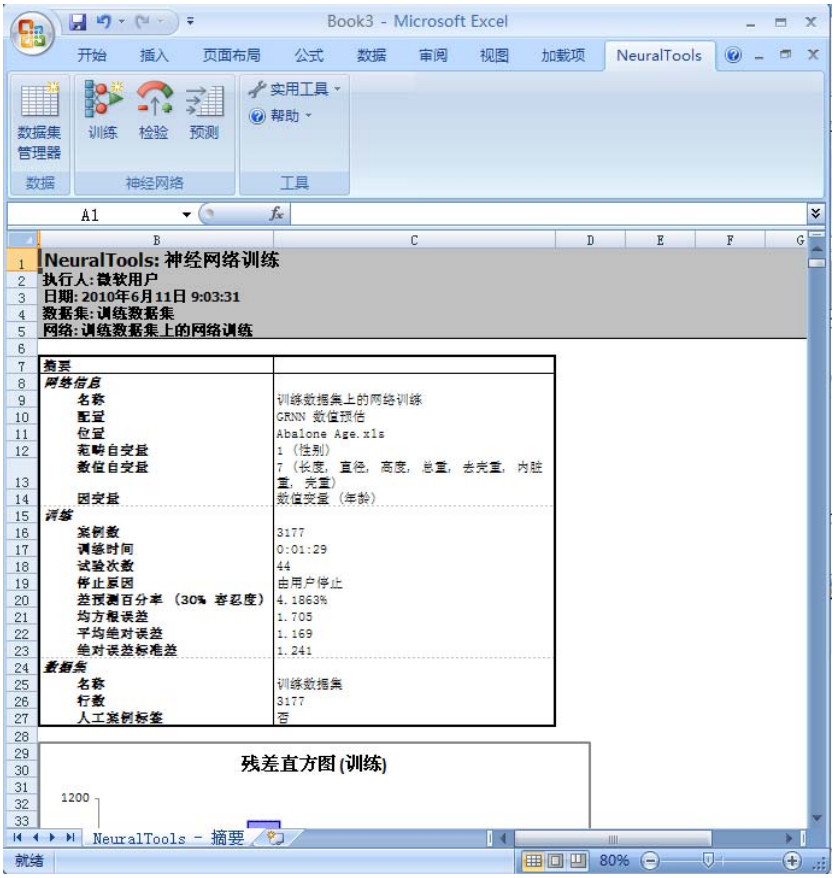
训练进度窗口报告训练数据的误差。此值的变化不足以得出有关网络对训练中没有使用的案例进行预测的质量情况的任何直接结论。这种结论应该根据检验数据的误差得出。另请注意，对于数值预测，进度窗口中报告的误差是基于缩放数据（请参阅『输入项变换』一节中的缩放信息）的均方根误差。对于范畴预测，根据范畴数据的数字表示来报告误差。



训练报表

训练后可创建摘要报表和详细报表。这些报表详细说明已训练的神经网络的性能。生成的实际报表内容在“应用程序设置”对话框中的“要生成的报表”和“详细报表中的列数”设置下指定。

- **训练摘要报表** - 训练摘要报表提供已训练的神经网络的性能统计量和图表。


Book3 - Microsoft Excel
开始
插入
页面布局
公式
数据
审阅
视图
加载项
NeuralTools
数据集管理器
训练
检验
预测
实用工具
帮助
数据
神经网络
工具
A1
NeuralTools: 神经网络训练
执行人: 微软用户
日期: 2010年6月11日 9:03:31
数据集: 训练数据集
网络: 训练数据集上的网络训练
摘要
网络信息
名称 训练数据集上的网络训练
配置 GRNN 数值预估
位置 Abalone Age.xls
范畴自变量 1 (性别)
数值自变量 7 (长度, 直径, 高度, 总重, 去壳重, 内脏重, 壳重)
因变量 数值变量 (年龄)
训练
案例数 3177
训练时间 0:01:29
试验次数 44
停止原因 由用户停止
差预测百分率 (30% 容忍度) 4.1863%
均方根误差 1.705
平均绝对误差 1.169
绝对误差标准差 1.241
数据集
名称 训练数据集
行数 3177
人工案例标签 否
残差直方图 (训练)
1200
NeuralTools - 摘要
就绪
80%

对于范畴预测/分类，训练摘要报表中的关键统计量和图表包括：

1) **不良预测百分率** - 预测范畴中与实际范畴不一致的案例百分比。

2) **平均不正确概率（仅对 PNN 网络提供）** - 对于每个案例，NeuralTools 都会计算不正确范畴的概率，它是 PNN 网络对不正确范畴的总概率所赋的值。例如，对于某个给定案例，网络对红色所赋的概率值为 30 %、对黄色为 20%、对绿色为 50%，且我们知道正确答案是红色，那么该案例的平均不正确概率值为 20% + 50% = 70%。对于范畴预测，此值提供一种因案例而异的误差量度，相当于数值预测的残差。“平均不正确概率”是所有案例的平均误差值。

   “详细报表”根据各个案例显示“不正确概率”。要更好地理解这一概念，可将“详细报表”的设置更改为显示概率神经网络对因变量的每个可能范畴所赋的概率值。要执行此操作，在**实用工具**菜单中选择**应用程序设置**，并单击**详细报表中的列数**一行右侧的下拉菜单。将显示 **NeuralTools - 要在详细报表中显示的列数**对话框。在该对话框中，选择**所有范畴的概率（针对 PNN）**以进行检验。然后选择**自动检验**，使用至少有三个因变量范畴的数据集训练 PN 网络（可使用 Auto Loans.xls 示例文件）。在生成的“详细报表”中，观察**不正确百分率**这一列的值与赋给每个可能范畴的概率有什么关联。不正确百分率是所有不正确范畴的概率总和。

3) **分类矩阵** - 逐个比较实际范畴和预测范畴。例如：分类矩阵可能显示一个网络正确地检测到患者所患的病症，但也有有可能对健康患者发出假警报。

4) **变量作用** – （如选择）显示自变量对预测答案的相对作用。

5) **不正确范畴的概率直方图**（仅向 PNN 网络提供） - 请参阅上述“平均不正确概率”以了解对“不正确范畴的概率”的说明。

对于数值预测，训练摘要报表中的关键统计量和图表包括：

1) **不良预测百分率** - 如果预测超出实际值的定义范围，则该预测计为“不良”；范围幅度在“应用程序设置”对话框的“良好/差容忍度(训练)”设置中定义。

2) **均方根误差** - 一种衡量预测值与实际值的偏差的指标（计算为平均方差的平方根）。

3) **平均绝对误差** - 预测值与实际值的平均偏差。

4) **变量作用** – （如选择）显示自变量对预测答案的相对作用。

5) **残差直方图** - “残差”是实际值与预测值的差。

6) **散点图**显示实际值、预测值和残差之间的关系。

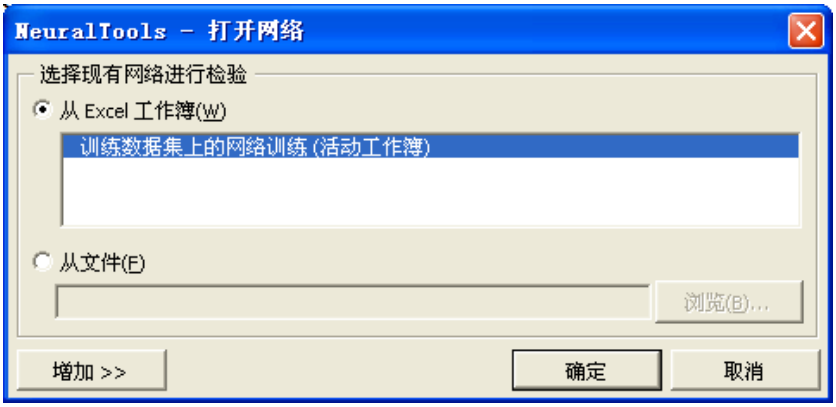
检验命令

指定用以检验一个已训练神经网络的设置，并进行检验

可使用**检验**命令 1) 指定用于检验一个已训练神经网络的设置，然后 2) 开始检验。

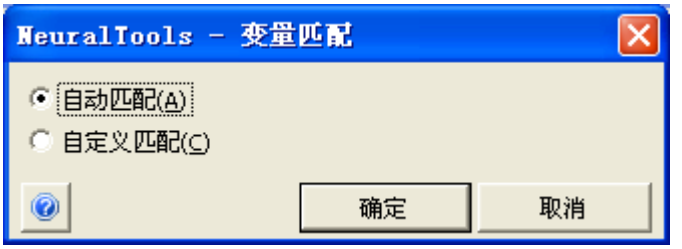
检验数据通常是输出值已知、但没有在训练网络时使用的数据。“检验”对话框中的选项包括：

- **数据集** – 显示将在检验已训练的神经网络时使用的数据集。此数据集需要使用数据集管理器进行定义并存在于当前工作表中。
- **要使用的网络** – 指定要检验的神经网络的名称和位置。可将神经网络保存到一个 Excel 工作簿中，也可保存到硬盘上的一个文件中。单击**浏览…** 以更改显示的名称和位置。



变量匹配

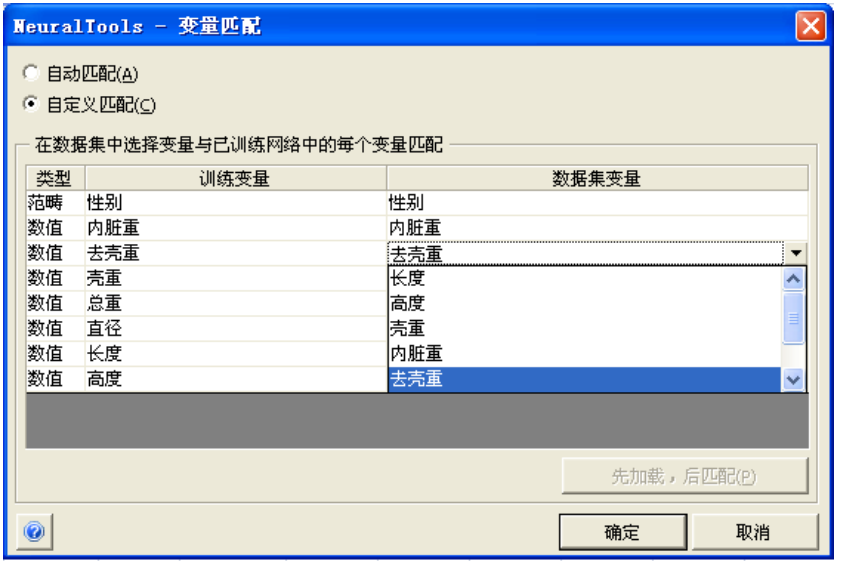
变量匹配指定数据集内要检验的变量与该数据集内用于训练网络的变量如何匹配。



可使用两个选项进行变量匹配：

1) **自动匹配**。检验数据集内的“变量名”与已训练的网络中数据集的变量名匹配，变量类型则根据这种匹配进行设置

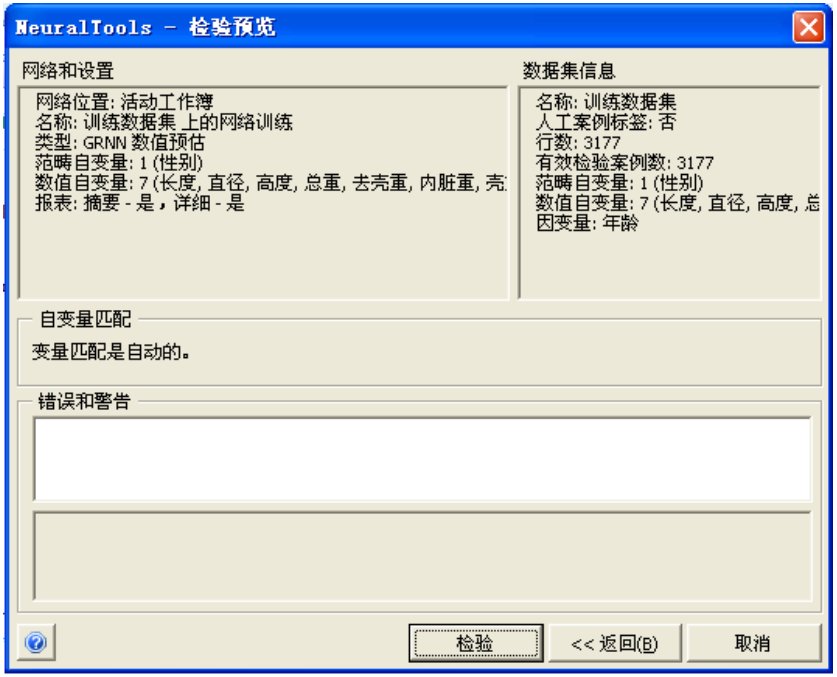
2) **自定义匹配**。可使用自定义匹配单独将检验数据集内的变量和已训练的网络中数据集的变量相匹配。仅当两个数据集中的变量名不同，或者需要不同赋值的时候，才可完成这种匹配。



“变量匹配”对话框列出每个数据集中的变量名，以便进行匹配。仅可匹配类型相同的变量。每次进行匹配时，所赋的值就会存储在数据集中。通过单击**先加载，后匹配**，您可浏览先前进行的匹配，以访问该数据集一组先前的赋值。

检验预览对话框

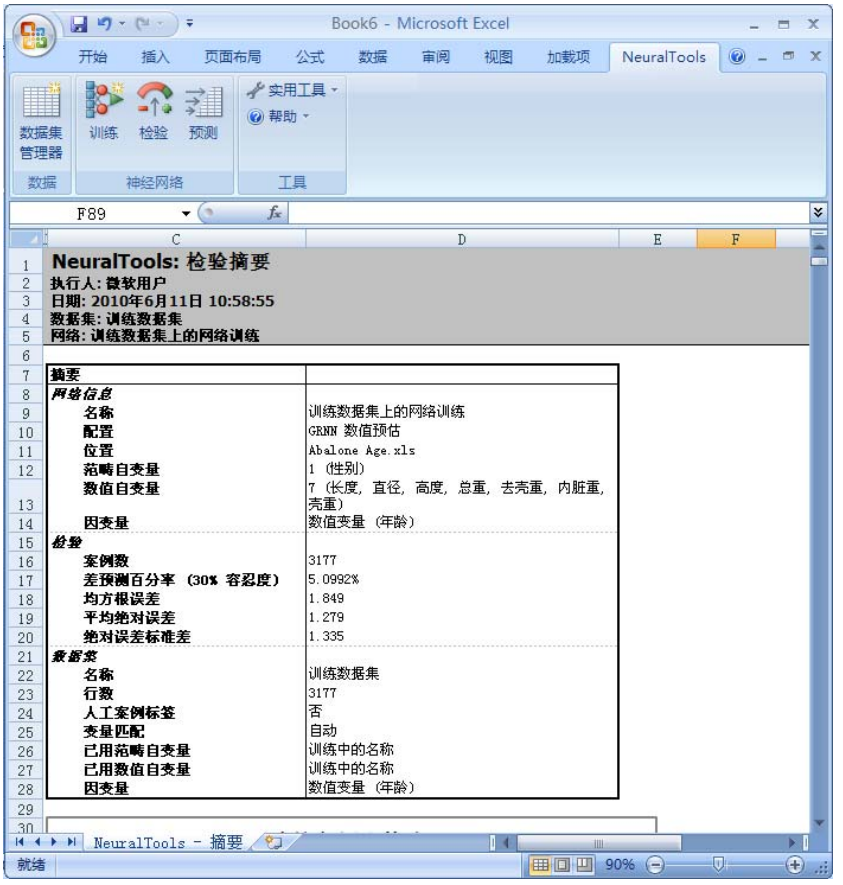
“检验预览”对话框在开始检验之前显示在当前的网络检验设置和数据中发现的错误。通过查看这一对话框的内容，您可根据 NeuralTools 的报告了解选择的所有检验假设。**错误和警告**部分描述了 NeuralTools 在数据中发现的问题，如有必要，您可在检验前进行更正。



检验报表

在检验后可创建摘要报表和详细报表。这些报表详细说明了已训练的神经网络在检验数据集方面的性能。生成的实际报表内容在**应用程序设置**对话框中的**要生成的报表**和**详细报表中的列数**设置下指定。检验时，使用详细报表特别有用，因为它显示了已训练的网络如何预测检验数据集中的各个输出值。

- **检验摘要报表** – 检验摘要报表提供已训练的神经网络在检验数据集方面的性能统计量和图表。


Book6 - Microsoft Excel
开始
插入
页面布局
公式
数据
审阅
视图
加载项
NeuralTools
数据集管理器
训练
检验
预测
实用工具
帮助
数据
神经网络
工具
F89
NeuralTools: 检验摘要
执行人: 微软用户
日期: 2010年6月11日 10:58:55
数据集: 训练数据集
网络: 训练数据集上的网络训练
摘要
网络信息
名称 训练数据集上的网络训练
配置 GRNN 数值预估
位置 Abalone Age.xls
范畴自变量 1 (性别)
数值自变量 7 (长度, 直径, 高度, 总重, 去壳重, 内脏重, 壳重)
因变量 数值变量 (年龄)
检验
案例数 3177
差预测百分率 (30% 容忍度) 5.0992%
均方根误差 1.849
平均绝对误差 1.279
绝对误差标准差 1.335
数据集
名称 训练数据集
行数 3177
人工案例标签 否
变量匹配 自动
已用范畴自变量 训练中的名称
已用数值自变量 训练中的名称
因变量 数值变量 (年龄)
NeuralTools - 摘要
就绪
90%

对于范畴预测，检验摘要报表中的关键统计量和图表包括：

1) **不良预测百分率** - 预测范畴中与实际范畴不一致的案例百分比。

2) **平均不正确概率（仅对 PNN 网络提供）** - 对于每个案例，NeuralTools 都会计算不正确范畴的概率，它是 PNN 网络对不正确范畴的总概率所赋的值。例如，对于某个给定案例，网络对红色所赋的概率值为 30 %、对黄色为 20%、对绿色为 50%，且我们知道正确答案是红色，那么该案例的平均不正确概率值为 20% + 50% = 70%。对于范畴预测，此值提供一种因案例而异的误差标准，相当于数值预测的残差。“平均不正确概率”是所有案例的平均误差值。

   “详细报表”根据各个案例显示“不正确概率”。要更好地理解这一概念，可将“详细报表”的设置更改为显示概率神经网络对因变量的每个可能范畴所赋的概率值。要执行此操作，在**实用工具**菜单中选择**应用程序设置**，并单击**详细报表中的列数**一行右侧的下拉菜单。将显示**NeuralTools - 要在详细报表中显示的列数**对话框。在该对话框中，选择**所有范畴的概率（针对 PNN）**以进行检验。然后选择**自动检验**，使用至少有三个因变量范畴的数据集训练 PN 网络（可使用 Auto Loans.xls 示例文件）。在生成的“详细报表”中，观察**不正确百分率**这一列的值与赋给每个可能范畴的概率有什么关联。不正确百分率是所有不正确范畴的概率总和。

3) **分类矩阵** - 逐个比较实际范畴和预测范畴。例如：分类矩阵可能显示一个网络正确地检测到患者所患的病症，但也有可能对健康患者发出假警报。

4) **不正确范畴的概率直方图**（仅向 PNN 网络提供） - 请参阅上述“平均不正确概率”以了解对“不正确范畴的概率”的说明。

对于数值预测，检验摘要报表中的关键统计量和图表包括：

1) **不良预测百分率** - 如果预测超出实际值的定义范围，则该预测计为“不良”；范围幅度在“应用程序设置”对话框的“良好/差容忍度(检验)”设置中定义。

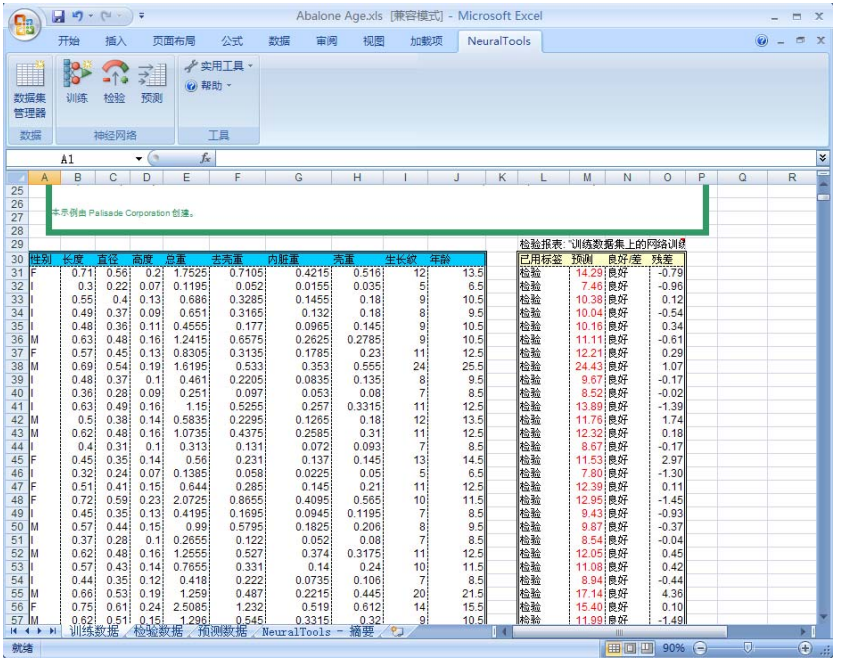
2) **均方根误差** - 一种衡量预测值与实际值的偏差的指标（计算为平均方差的平方根）。

3) **平均绝对误差** - 预测值与实际值的平均偏差。

4) **残差直方图** - “残差”是实际值与预测值的差

5) **散点图**显示实际值、预测值和残差之间的关系。

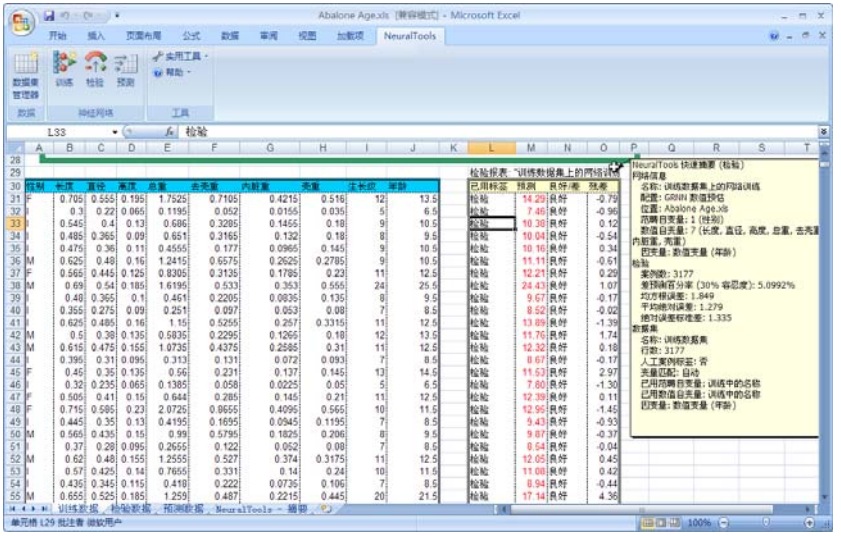
- **检验详细报表。**此报表位于检验数据集旁边，显示已训练的网络如何预测检验数据集中的各个输出值。



在“检验详细报表”中，根据“应用程序设置”对话框中的容忍度设置，对各项预测标示为**“良好”**或**“不良”**。如果运行多个检验，会在检验数据集右侧的新列中添加“详细报表”，以便您了解检验已训练的新网络时，对每个案例的预测有何变化。

详细报表中的快速摘要

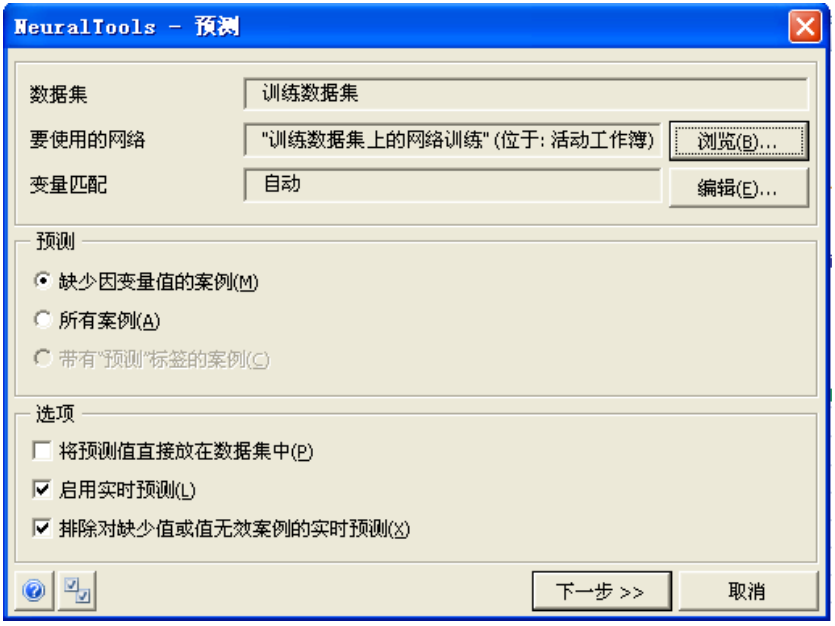
查看“详细报表”时，可使用 Excel 中的弹出注释快速访问“摘要报表”信息。只要将鼠标移到报表头，就会显示该弹出注释。注意：必须在 **Excel 工具命令选项对话框视图选项卡**中启用注释，才能显示弹出注释。



预测命令

指定使用已训练的神经网络预测值的设置，并进行预测

可使用**预测**命令来：1) 指定使用已训练的神经网络预测值时的设置，以及 2) 进行预测。



要预测的数据通常是因变量值未知的数据。“预测”对话框中的选项包括：

- **数据集** – 显示要用于预测的数据集。此数据集需要使用数据集管理器进行定义并位于当前工作表中。
- **要使用的网络** – 指定要用于预测的神经网络的名称和位置。可将神经网络保存到一个 Excel 工作簿中，也可保存到硬盘上的一个文件中。单击**浏览…** 以更改所显示的名称和位置。
- **变量匹配** – 指定数据集中有预测数据的变量与该数据集中用于训练网络的变量如何匹配。单击**编辑…**以更改变量匹配。有关**变量匹配**的更多信息，请参阅本章中的**检验**命令。
- **预测** – 选择将进行预测的案例。通常，您会选择**缺失因变量值**的案例进行预测，不过如果愿意，也可以对**所有案例**（甚至是因变量值已知的案例）进行预测。如果您的数据集中有一个标签变量，则只能对标有“预测”标签的案例预测因变量值。

- **选项** – 设置预测值的位置和实时预测选项。

  1) **将预测值直接放在数据集中。**此选项指定预测值将直接放在数据集中每个预测案例的因变量位置，可能还会放在“详细报表”中（这取决于是否在“应用程序设置”中的“要生成的报表”下选择了“详细报表”）。这会覆盖单元的当前内容，因此应谨慎使用。您可通过数据集内的颜色来辨别预测值。

  2) **启用实时预测** - 指定 NeuralTools 会将公式放在显示因变量预测值的单元中。这些公式可让 NeuralTools 根据自变量值的变化来计算预测值。

  3) **排除对缺失值或无效值案例的实时预测** – 指定在案例缺失输入变量值的情况下，不会添加实时预测公式。缺失输入值会导致实时预测公式返回一个错误值。不过，允许 NeuralTools 对缺失自变量值的案例输入公式，这种情况公式可以使用，因为只要填充缺失的值，就会自动显示预测。

实时预测

实时预测是 NeuralTools 的一个强大功能（仅向正式版提供），使您无需进行具体的“预测”操作，即可在 Excel 中自动执行预测。进行实时预测时，NeuralTools 会将公式放在显示因变量预测值的单元中。这些公式使用一个自定义 NeuralTools 函数来计算预测值，如：

*=**NetOutputPrediction**(_PALDS_DG25B8C82B!$A$140, "DG25B8C82B", "VG1DD83AF2", 'Prediction Data'!$A$6:$I$6, A7:I7)*

实际公式由 NeuralTools 添加到您的工作表中，不需要由您来输入。这些引数可让 NeuralTools 确定正在使用的已训练网络，以及工作表中自变量值的位置。当添加或更改某案例的输入自变量值时，NeuralTools 会自动返回一个新的预测值。这样，使用现有的已训练网络为新案例添加或生成预测变得简单易行。

注：如果预测基于不希望被更改的单元值，则建议在训练或预测对话框中取消选择“实时预测”；这样会尽量减少 Excel 重新计算工作簿的时间。

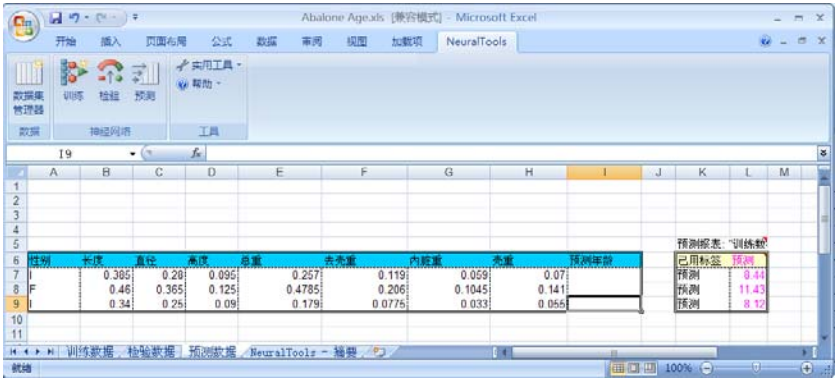
预测预览对话框

“预测预览”对话框在开始预测前显示选定数据集的预测设置，以及在数据或设置中发现的错误。通过查看这一对话框的内容，您可根据 NeuralTools 的报告了解选择的所有预测假设。**错误和警告**部分描述了 NeuralTools 在数据中发现的问题，如有必要，您可在预测之前进行更正。

预测报表

预测后可创建摘要报表和详细报表。这些报表详细说明了已训练的神经网络在检验数据集方面的性能。生成的实际报表内容在**应用程序设置**对话框中的**要生成的报表**和**详细报表中的列数**设置下指定。

- **详细预测报表。**此报表位于预测数据集旁边。它提供了一个预测值存放位置，以便用户在不想将预测值放在数据集本身的因变量内时使用；如果某些案例的因变量包含历史数据，不要将历史案例与网络预测混在一起，这样会比较安全。

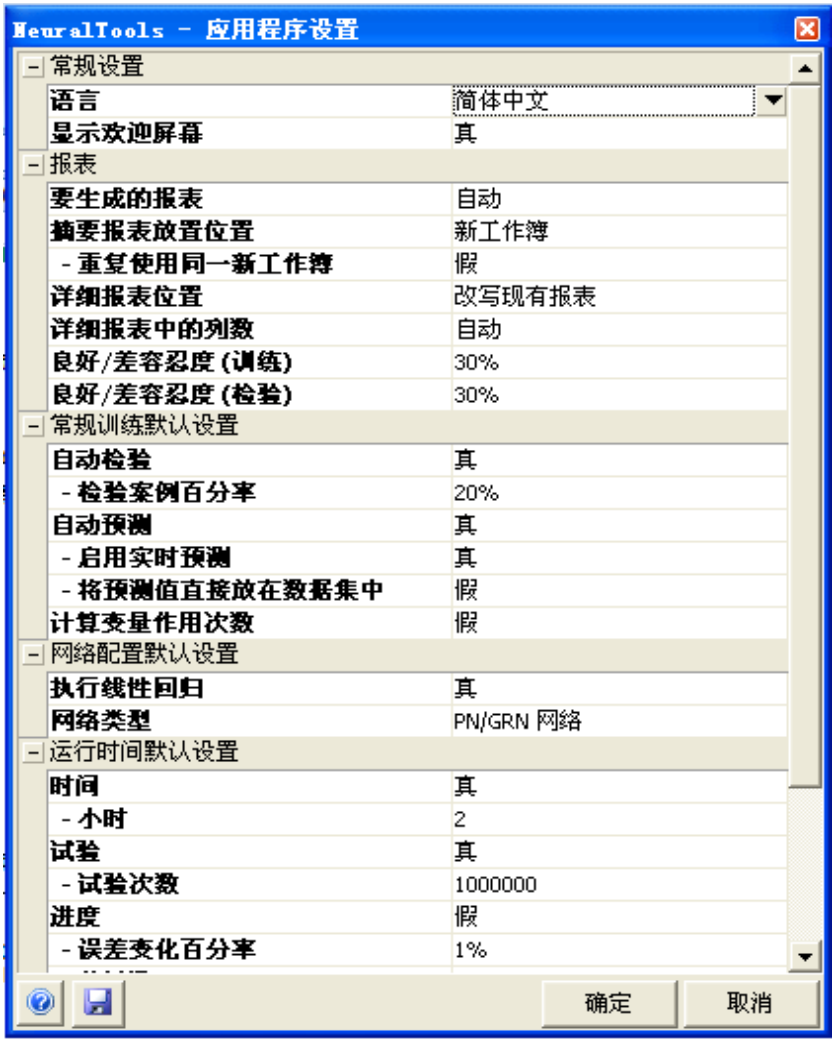


如果运行多个预测，会在数据集右侧的新列中添加“详细报表”，以便您了解使用已训练的新网络时，对各个案例的预测有何变化。

应用程序设置命令

指定训练、检验和预测的默认设置。

可使用**应用程序设置**命令来选择 1) 对训练、检验和预测生成什么报表，2) 训练默认使用的选项以及 3) 预测和运行时间默认使用的选项。在训练、检验或预测对话框中，默认使用许多应用程序设置。请参阅对这些对话框的说明，以了解这些设置的更多信息。其他设置如下所示。



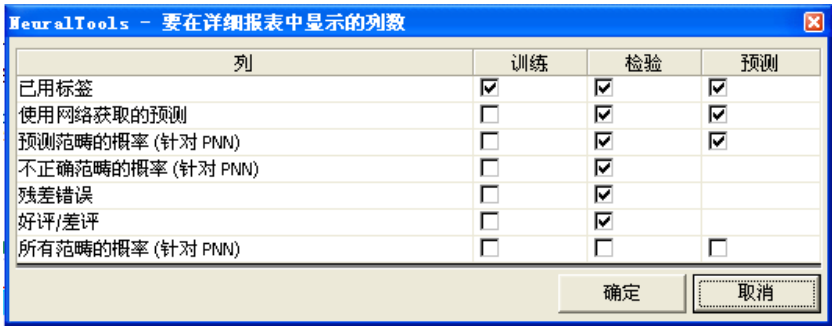
报表设置包括：

- **要生成的报表** - NeuralTools 中的每个操作均可生成摘要报表和详细报表。但是，通常您会希望使用默认报表设置，因为某些报表对某些操作的价值不大。例如，“详细报表”是预测操作的标准报表，而在这种情况下，“摘要报表”没有多大价值。

“摘要报表”自己有一个工作表，而“详细报表”则放在数据集右侧的列中，与数据集同在一个工作表中。

- **摘要报表放置位置**选项包括：
  - **新工作表**，在此可为每份报表创建一个新工作表。您可以对各个报表重复使用同一个工作簿，也可每次都创建一个新工作簿。
- **详细报表位置**选项包括：
  - **改写现有报表**，该选项会使用新的“详细报表”改写数据集中包含先前“详细报表”数据的列（要手工删除一个“详细报表”，单击并拖动选择列标题，选中包含报表数据的所有列，然后从“编辑”菜单中选择“删除”）。
  - **数据集的右侧**，该选项会将新列插入数据集的右侧，以放置新的“详细报表”。
  - **现有报表的右侧**，该选项会使用数据集右侧的列和现有报表来存放新的“详细报表”。

- **要在详细报表中显示的列数。**对于选中的每行，数据集右侧的“详细报表”中会显示一列，给出每个案例的信息。



可显示以下列：

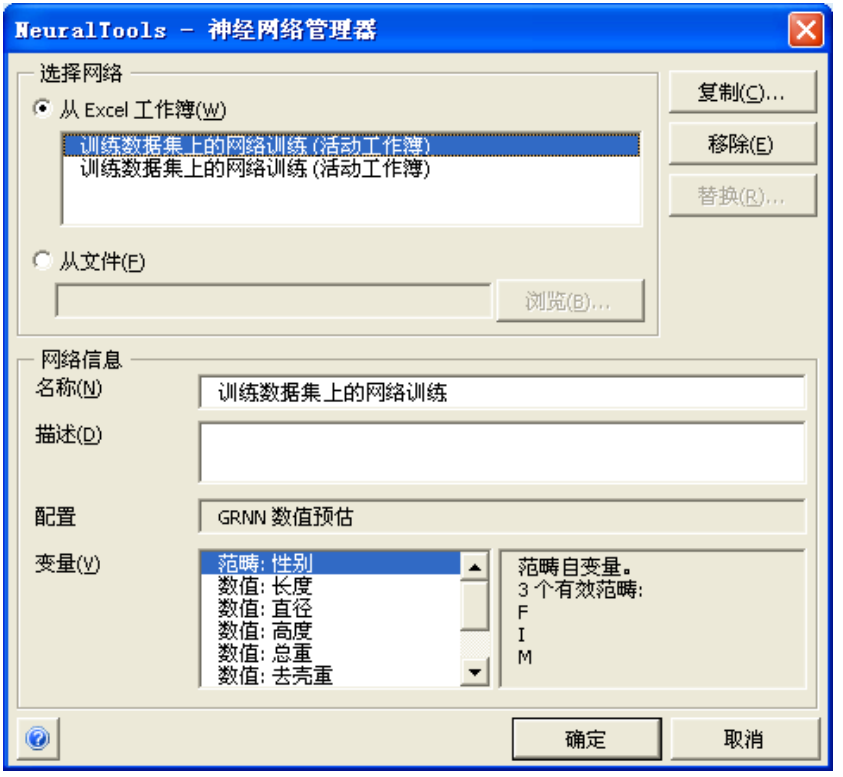
1) **已用标签 -** 对于训练集或检验集当中的每个案例，显示“训练”或“检验”标签，对于进行了预测的给定案例，显示“预测”标签。
2) **使用网络获取的预测 -** 网络预测的数值或范畴。
3) **预测范畴的概率（针对 PNN） -** 概率神经网络不仅可预测未知范畴，还可对该范畴的概率赋值。使用多层前馈网络对范畴进行预测时不可用。不适用于数值预测。
4) **不正确范畴的概率（针对 PNN） -** PNN 网络对不正确范畴所赋的概率值。例如，对于某个给定案例，网络对红色所赋的概率值为 30 %、对黄色为 20%、对绿色为 50%，且我们知道正确答案是红色，那么该案例的平均不正确概率值为 20% + 50% = 70%。对于范畴预测，此列提供一种因案例而异的误差标准，相当于数值预测的“残差”列。
5) **残差 -** 实际因变量值与预测因变量值之差。不适用于范畴预测。
6) **好评/差评 -** 对于数值预测，此列说明对给定案例的预测是否在实际值上下的定义范围内。范围幅度定义为“好评/差评容忍度”。对于范畴预测，此列只是说明预测范畴与实际范畴是否一致。
7) **所有范畴的概率（针对 PNN） -** 如果选择了此选项并训练了概率神经网络，将对所有的范畴因变量插入一列。例如，网络用于预测颜色，可能有“red%”、“yellow%”和“green%”这些列，其中包含对每种颜色所赋的概率值。

- **好评/差评容忍度。**对于检验和训练，如果数值预测在输入的实际因变量值百分比内，则会标为“良好”。

神经网络管理器命令

允许复制、移动和删除已训练的神经网络

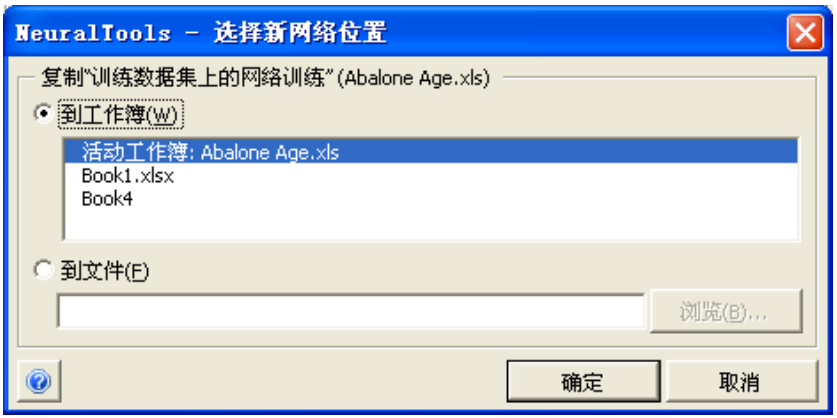
可使用神经网络管理器命令管理已训练的神经网络，将其在工作簿和文件之间移动以及向其添加描述信息。



可将神经网络保存到一个 Excel 工作簿中，也可保存到硬盘上的一个文件中。可以在一个 Excel 工作簿中放置任意数量的网络。使用神经网络管理器，您可将网络移到新的工作簿或文件中，或者删除、替换这些网络。这样，无需附带训练数据的工作簿，您就可以方便地使用已培训的现有网络分析其他工作簿中的数据集。

神经网络管理器选项包括：

- **复制** – 该选项将已训练的神经网络复制到其他位置。只需选择您想放置网络的工作簿或文件即可。



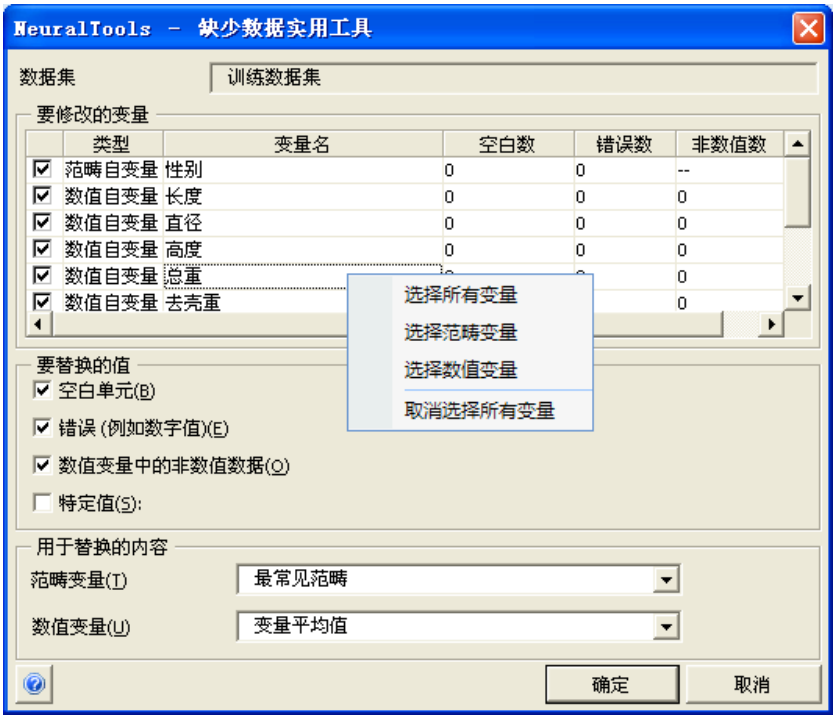
- **移除** – 删除已训练的神经网络。
- **替换** – 用新的神经网络覆盖已训练的神经网络。对用于“实时预测”的网络提供此功能。替换后，先前使用旧网络进行的实时预测将改用新网络进行预测。然而，这并不适用于“详细报表”。如果详细报表包含的实时预测单元中使用了要被替换的网络，替换后，这些单元将包含固定值。
- **网络信息** – 可向网络添加描述信息。这有助于识别已训练的网络及其训练条件。

缺失数据实用工具命令

可用人为值替换数据集内缺失的数据值和错误值。

可使用“缺失数据实用工具”命令用人为值替换数据集内缺失的数据或其它不需要的数据。在训练、检验和预测时，NeuralTools 会忽略缺失变量值的案例，因此在处理前更正这些错误通常很有用。

如果您的数据集中有缺失值的案例，“训练预览”对话框将给出警告。如果出现这种情况，可使用“缺失数据实用工具”命令进行更正。



“缺失数据实用工具”对话框有以下选项：

- **要修改的变量** - 提供当前工作表中的数据集使用的一组变量并显示缺失数值、错误以及（对于数值变量）非数值数据。勾选一个变量，以替换其缺失的数据或其它不需要的数据。

  变量表提供了一个右键单击菜单，其中包含选择和取消选择成组变量的命令。

- **要替换的值** – 为选中要替换的变量选择值的类型。**特定值**可用于将某个变量的所有特定值替换为一个新值。
- **用于替换的内容** – 在数据集中指定用来替换缺失数据或其它不需要数据的值。对于范畴变量和数值变量，指定的值有所不同：

  范畴变量 – 选项有：

  - **最常见或不常见范畴** – 数据集内的案例最常见或最不常见的范畴值。
  - **相邻范畴** – 数据集中与缺失值的案例相邻的案例所发生的范畴值。
  - **随机选择的范畴** – 从数据集的范畴值中随机选择的范畴值。
  - **特定范畴** – 将所有缺失的或不需要的值设为一个特定值。

  数值变量 – 选项有：

  - **变量平均值** – 变量在数据集的所有案例中的平均值
  - **变量中位数值** – 变量在数据集的所有案例中的中位数值
  - **来自相邻值的插值** – 对于数据集中缺失值的案例，在相邻案例的变量值之间通过内插计算得出的值
  - **随机值（介于最小值和最大值之间）** – 对数据集内所有案例选择的介于变量最小值和最大值之间的随机值

对于以上两种变量类型，均可使用**清除单元**来清除为变量选择的值。

有关缺失值的更多信息

“缺失数据实用工具”对话框为缺失的数据提供一个可能的解决途径：它在缺失实际数据的位置生成人为数据。不过使缺失的数据保留为空白单元，让 NeuralTools 忽略缺失数据的案例会更好。请注意，NeuralTools 不会将诸如 "?" 之类的特殊符号识别为缺失数据；需要清除问号，此步骤可通过“缺失数据实用工具”完成，方法是在“要替换的值”部分选择“特殊值”，并在“用于替换的内容”部分选择“清除单元”。

还可以借助很少或没有缺失数据的自变量，使用 NeuralTools 来预测一个自变量的缺失值。检验结果会表明训练用来预测缺失值的网络是否可靠。

神经网络的更多信息

神经网络基本知识

神经网络是输入数值、并对这些输入项进行计算，然后输出一个或多个数值的系统。当针对某种应用设计并训练一个神经网络之后，其输出逼近给定输入项的正确值。例如：一个网络有若干输入项，代表鲍鱼（一种海洋动物）的一些易于衡量的特征，如：长度、直径和重量。网络中执行的计算会得出一个数值，通常接近该动物的年龄（鲍鱼的年龄较难确定）。

神经网络这种构思源于大脑的结构。大脑由大量单元组成，这些单元称为“神经元”。一个神经元通过若干“树突”接收其他神经元的刺激。神经元根据接收的刺激，通过其唯一的“轴突”向其他神经元发送信号，这些“轴突”与其他神经元的“树突”相连。人工神经网络与大脑一样由很多元素组成，每个元素都接收大量输入，生成一个输出，而该输出是输入的一个相对简单函数。

神经网络与统计方法

神经网络提供了一种取代更为传统的统计方法的途径。它们与线性回归一样，用于函数逼近。还与“判别分析”和“逻辑回归”一样，用于分类。神经网络的优势是能对极其复杂的函数建模。这与传统的线性技术（线性回归和线性判别分析）截然不同。在 20 世纪中期，人工神经网络发明之前，优化线性模型的技术就广为人知了。训练神经网络的有效算法经过多年才得以建立。但是，我们如今已有了一系列训练神经网络的先进算法，使该网络成为一种颇有吸引力的取代传统方法的选择。

神经网络的结构

神经网络的结构是由相互连接的单元构成的，这些单元称为**“节点”**或**“神经元”**。每个神经元执行网络内的一部分计算：一个神经元将一些数字作为输入项，对这些输入项执行相对简单的计算，然后返回一个输出项。一个神经元的输出值传送到另一个神经元，作为后者的输入项，而生成整个系统的最终输出值的神经元不再传送值。

神经元按**层**排列。输入层神经元接收输入项（如：一个鲍鱼的长度、直径和重量）以进行计算。这些值传递给第一个隐含层的神经元，后者对这些输入项进行计算，然后将其输出项传到下一层。如果有下一层，它可能还是一个隐含层。最后一个隐含层中各神经元的输出项传递给生成网络最终输出项（如：鲍鱼的年龄）的神经元。

数值预测和范畴预测

将神经网络用于预测数值时，它们通常只有一个输出项。这是因为单输出项网络比多输出项网络更可靠，几乎所有预测问题都可使用单输出网络来加以解决。例如：对于未来几天的股票交易量和股价，最好构建一个网络来预测股价，构建另外一个网络来预测股票交易量，而不是构建一个既预测股票交易量又预测股价的网络。另外，在用于分类/范畴预测时，神经网络有多个输出。例如，假设我们想预测某只股票的价格在次日会“上升 1% 以上”、“下降 1% 以上”还是“变化不超过 1%”。这样，网络将有三个输出数值，最大的输出值就代表网络选择的范畴。

训练网络

训练网络是微调计算参数的过程，旨在使网络输出逼近给定输入项的正确值。这一过程一方面是由训练数据引导的，另一方面是由训练算法引导的。训练算法选择若干组计算参数，通过对每个训练案例应用网络来评估每组参数，以确定网络给出答案的理想程度。每组参数都是一个“试验”；训练算法根据之前的试验结果选择若干组新参数。

神经网络的计算机处理

神经网络是一种可在各种计算机硬件上执行的计算模型。可使用若干小处理部件建立一个神经网络，每个部件执行一个神经元的工作。不过，神经网络通常在有一个强大处理器的计算机上执行，即当今使用的大多数计算机。采用单处理器计算机，诸如 NeuralTools 之类的程序使用一个处理器来执行所有神经元的计算；在这种情况下，神经元的概念表述的是获得预测所需的那部分计算，与实际处理部件不同。

神经网络的类型

各种神经网络在结构、神经元内执行的计算以及训练算法上各有不同。NeuralTools 提供的一种神经网络是**多层前馈网络** (MLF)。通过 MLF 网络，NeuralTools 用户可指定需要一层还是两层隐含神经元，以及隐含层中应包含多少神经元（NeuralTools 可帮助您做出适当的选择，请参阅 MLF 网络部分的说明）。NeuralTools 还提供**广义回归神经网络** (GRN) 和**概率神经网络** (PN)；二者密切相关，前者用于数值预测，后者用于范畴预测/分类。使用 GRN/PN 网络，用户无需确定网络的结构。这些网络总是有两层隐含神经元，在第一个隐含层，每个训练案例都有一个神经元，第二层的大小取决于训练数据的某些因素。

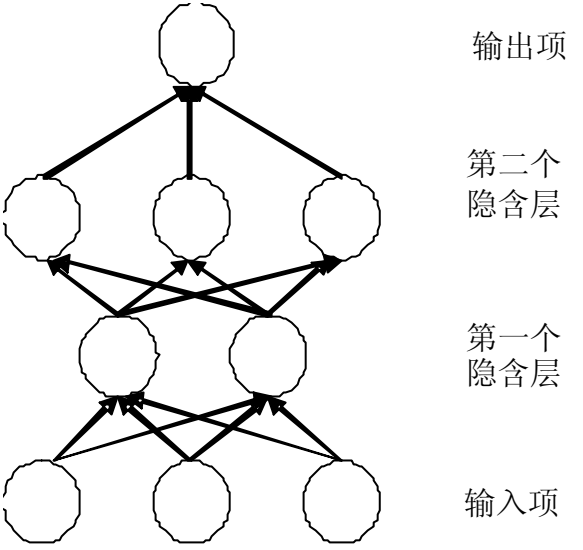
本章的以下部分将详细探讨 NeuralTools 提供的每一种神经网络。

多层前馈网络

多层前馈网络（也称为“多层预测网络”）是能逼近复杂函数，进而能对自变量和因变量之间的复杂关系建模的系统。

MLF 结构

下图显示了使用三个数值自变量进行数值预测的 MLF 网络；该网络的配置为第一个隐含层有 2 个神经元/节点，第二个隐含层有 3 个神经元/节点。



网络的行为由以下因素确定：

- 其拓扑结构（隐含层数量和这些层中的节点数）
- 连接的“权重”（分配给每个连接的参数）和偏差项（分配给每个神经元的参数）
- 激活/传递函数，用于将每个神经元的输入项转化为输出项

尤其应注意，有 N 个输入项的隐含神经元先计算其输入项的加权和：

$$和 = in_0 * w_0 + in_1 * w_1 + ... + in_n * w_n + 偏差项$$

其中，in_0 到 in_n 是上一层神经元的输出项，而 w0 到 wn 是连接权重；每个神经元都有自己的偏差值。然后，对该*和*应用激活函数以生成神经元的输出项。

S 形函数用作隐含层神经元的激活函数。尤其值得一提的是，NeuralTools 使用双曲正切函数。在 NeuralTools 中，将输出神经元用作激活函数。即：它只返回其输入项的加权和。神经网络有时采用对输出神经元使用 S 形激活函数的构造。不过，神经网络不一定需要逼近复杂函数。此外，S 形函数有严格的输出项范围（对双曲正切函数为 -1 到 1），通常会有超出范围的因变量值。因此，在输出神经元中使用 S 形函数必须对输出值再进行一次变换，然后再将训练数据传递给网络。

使用 MLF 网络进行分类时，这些网络有多个输出神经元，每个神经元对应一个可能的范畴因变量。网络通过计算输出数值来对一个案例进行分类；所选范畴是与输出最高值的神经元对应的范畴。

训练 MLF 网络

训练一个 MLF 网络包括找出一组可以让网络对新案例给出正确答案的连接权重和偏差项（为了简单起见，以下陈述将省去偏差项）。训练由分配一组随机选择的连接权重入手，然后对每个训练案例进行预测（通过将自变量值作为输入项来获得输出项）。输出项很可能与已知因变量值不同。因此，对于每个训练案例，我们都有一个误差值。从这些值中计算出整个训练集的误差标准；它会告诉我们在给定的初始权重下，网络的表现如何。

网络在最初随机分配的权重下表现可能欠佳，我们会继续进行试验：使用其他权重分配。不过，分配的权重不再是随机的，而是由我们的训练算法确定的，即根据先前试验的结果选择连接权重的方法。问题是一个优化问题：我们期望通过更改连接权重尽量减小误差标准。

历史背景

在 MLF 网络中训练连接权重的第一个成功的算法是“反向传播”；研究人员目前趋向于采用其他算法，因为这些算法更快，也更可能找到全局最优值。NeuralTools 使用“多层感知器”法，它属于“二级”优化法范畴。这些“决定性的”优化法旨在找出一个函数的局部最小值，它们沿着误差函数的曲线有效递减。为了降低找到局部最小值而不是全局最小值的风险，NeuralTools 综合使用了“决定性”优化法和“随机性”优化法。尤其值得一提的是，随机“模拟退火”法与多层感知器法配合使用。该算法根据先前试验的结果，决定在某一点使用什么方法。有关多层感知器法的更多信息，请参阅 Bishop (1995) 和 Masters (1995)。有关模拟退火法的更多信息，请参阅 Masters (1995)。

误差标准

训练数值预测网络时使用的误差标准是所有训练案例的均方差，它是正确答案和网络给出的答案之间的平均方差。对于分类，每个训练案例有多个输出项（每个输出项对应一个范畴因变量）。我们参考期望输出值，计算所有训练案例的所有输出项的均方差：对于每个训练案例，我们希望输出值接近 1，即：输出值与正确的范畴对应；同时希望其余输出值接近 0。

训练时间

NeuralTools MLF 训练算法采用不同的起始权重多次进行重新计算。因此，训练网络的时间越长越好。等待重新计算的时间越长，就越有可能找到误差函数的全局最小值。

拓扑结构选择

对层数和各层中的神经元数目的选择决定网络是否能学习自变量和因变量之间的关系。通常有一个隐含层和两个隐含神经元的网络不能训练到满意的误差水平。但是增加层数和节点数又太不划算。一个隐含层足以应对几乎所有问题；使用两层通常会无谓地延长训练时间。而且，一个隐含层中有几个神经元通常就足够了。

NeuralTools 可以根据训练数据自动配置网络拓扑结构。但是，“最佳网络搜索”功能可提供更可靠的方法。进行最佳网络搜索时，会训练一组神经元数量各异的单隐含层网络。默认情况是纳入五个 MLF 网络，其中有 2 到 6 个隐含神经元。如果有足够的时间，还可扩大范围；不过建议从含 2 个神经元的网络入手，以防过度训练。

防止过度训练

术语“过度训练”所指的情况是网络不仅学习了自变量和因变量的一般关系特征，还开始学习了训练案例中通常不用的内容；即：这些内容不会用于训练中没有涵盖的案例。有时要解决这一问题，会将检验集分为训练时检验集和训练后使用的正式检验集。训练时会定期计算训练时检验集的误差。误差开始增加时，表明网络开始过度训练，并且训练会停止。

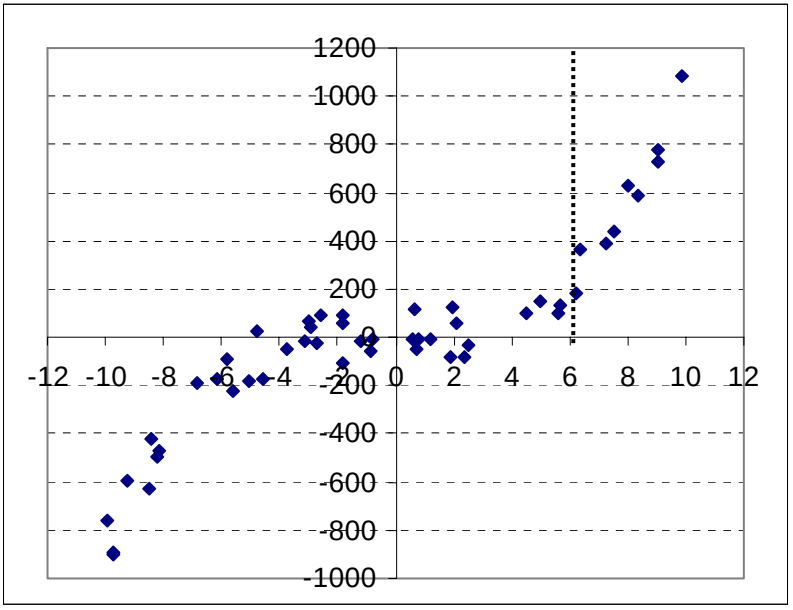
NeuralTools 采用另一种方式来防止过度训练。使用两个不同检验集的方式通常不现实，因为通常没有足够的数据分成一个训练集和两个检验集。另外，训练时检验集的误差增加也不是一个可靠的过度训练指标；增加可能是局部的，随着更多的训练，以后的误差可能会降低。NeuralTools 的最佳网络搜索旨在防止过度训练。在默认设置下，最佳网络搜索会从有 2 个神经元的网络开始，这种网络通常小到不可能过度训练。在默认设置下，该搜索将训练最多 6 个神经元的网络。如果有 5 个和 6 个神经元的网络过度训练，将显示在一个检验集的结果中；有 2、3 或 4 个神经元的网络检验误差最低。

广义回归神经网络和概率神经网络

广义回归神经网络和概率神经网络基于类似的理念。GRN 网络用于数值预测/函数逼近，而 PN 网络用于范畴预测/分类。两种网络均由 Donald Specht 提出 (参见 "Probabilistic Neural Networks", *Neural Networks*, 3, 1990, pp.109-118; "A General Regression Neural Network", *IEEE Transactions on Neural Networks*, 2, 1991, pp.568-576)。Masters (1995) 中也介绍了这两种网络，其表述总结如下。请参考这些资源以了解更多详情。

广义回归神经网络

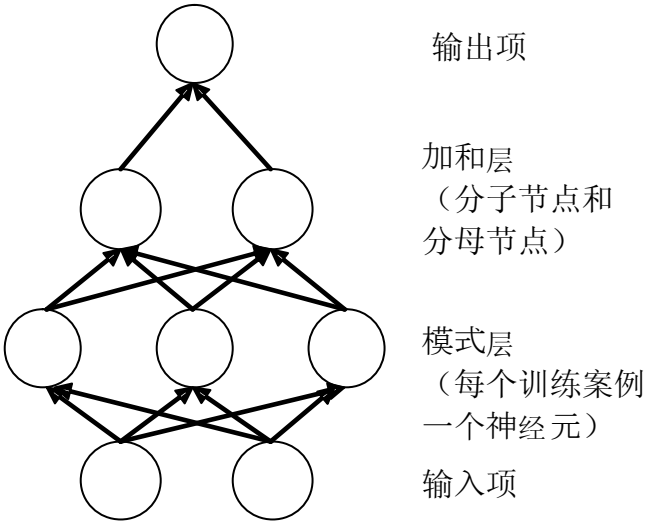
例如，请看图中的训练数据集，它有一个数值自变量和一个数值因变量。



观察人员可辨别数据的模式。如需解出自变量 6 对应的未知因变量值，可以估算到它大于 200 且小于 400。注意，这一估算并非根据两个最近的已知案例，它们指示值小于 200；我们参考的是其他案例。但是，我们没有注意自变量值在 -10 左右的案例；已知案例与未知案例越接近，在估计未知因变量时所占的权重就越高。广义回归神经网络就是建立在这些直观理念上的。网络中显示了每个训练案例。当网络遇到一个案例时，会使用每个训练案例的因变量值计算因变量预测值，越接近的训练案例对输出值的影响越大。

GRN 结构

有两个数值自变量的广义回归神经网络结构如图所示（假定只有三个训练案例）：



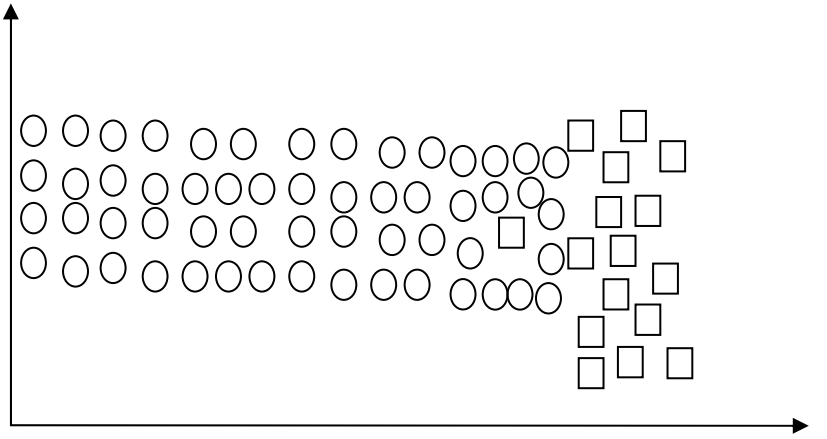
模式层包含每个训练案例的一个节点。本例中，向网络提供训练案例就是提供两个数值自变量。模式层中的每个神经元计算其与提供案例的距离。传递给分子节点和分母节点的值是距离和因变量值函数。加和层的两个节点加总其输入项，而输出节点则将其分开以生成预测值。

模式层神经元计算的距离函数使用“平滑因子”；每个输入项均有自己的 “平滑因子”值。在有一个输入项的情况下，平滑因子的值越大，训练案例离预测值的距离就越远。在有两个输入项的情况下，平滑因子沿一个平面上的一个轴与距离的相关，通常在有多个输入项的情况下，平滑因子与多维空间的一维相关。

训练 GRN 网络就是优化平滑因子，以尽量降低训练集的误差；使用多层感知器优化法来完成这一任务。训练时用来评估不同组平滑因子的误差标准是均方差。不过，在计算一个训练案例的方差时，会将该案例从模式层暂时排除。这是因为被排除的神经元会算作一个零距离，降低其他神经元在预测计算中的重要程度。

概率神经网络

现在讨论概率神经网络。请看以下训练数据集，它有两个数值自变量和一个有两个范畴的因变量：

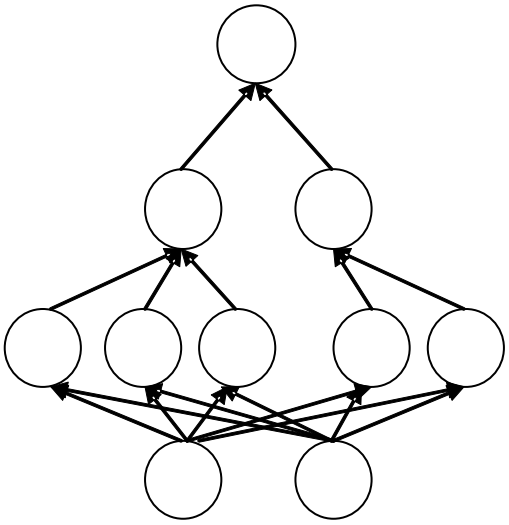


圆圈代表一个范畴中的训练案例，而方框则代表另一范畴的案例。我们想预测显示为问号的案例属于哪个范畴。观察人员会认为，该案例更有可能属于圆圈范畴，而不是方框范畴。但是，许多分类法却不能得出相同的结论。那些要求范畴有线性可分性的方法得不出结论。最近邻法会将未知案例分配给方框范畴。那些侧重集中趋势的方法也是如此，因为未知案例更接近方框范畴的质心，而不是圆圈范畴的质心。

而 PN 网络则会做出正确的预测。它将考虑新案例与每个训练案例的距离，对更接近的案例分配更高的权重。对于邻近的圆圈，旁边方框的影响会降低。

PNN 结构

概率神经网络的结构如图所示，该图假设有两个数值自变量、两个范畴因变量和五个训练案例（其中三个在一个范畴，两个在另一个范畴）；



向网络提供一个案例时，模式层中的每个神经元计算自己所代表的训练案例和输入案例之间的距离。传递给加和层神经元的值是距离和平滑因子的函数。与 GRN 网络一样，每个输入项本身都有一个平滑因子；这些因子确定训练案例的重要性随距离远近降低的速度。在加和层，每个范畴因变量都有一个神经元；每个神经元加总该范畴内的训练案例所对应的所有神经元的输出值。加和层神经元的输出值可视为每个类的概率密度函数预测。输出神经元选择概率密度函数值最高的范畴作为预测的范畴。

与 GRN 网络一样，训练一个 PN 网络就是优化平滑因子，以尽量降低训练集的误差；并使用多层感知器优化法。训练时用来评估不同组平滑因子的误差标准是根据加和层神经元返回的所有训练案例的所有值计算出来的。这种标准不仅考虑了正确范畴的概率，还考虑了不正确范畴的概率分布（概率在不正确范畴之间的分布大体一致，比某些不正确范畴的概率较高要好）。请注意，在计算一个训练案例的误差时，会将该案例从模式层暂时排除。这是因为被排除的神经元会算作一个零距离，降低其他神经元在计算中的重要程度。

MLF 网络与 PN/GRN 网络的比较

NeuralTools 中提供的每种神经网络都有各自的优缺点，具体如下：

GRN/PN 网络的优点

- 训练速度快
- 不需要指定拓扑结构（隐含层和隐含节点的数目）
- PN 网络不仅可以分类，还可以返回案例落在其他因变量范畴的概率

MLF 网络的优点：

- 较小，因此预测速度也较快
- 对于不在训练数据范围内的案例更可靠（例如：某个自变量的值不在训练数据中该变量值的范围内）；不过请注意，使用 MLF 网络进行不在训练数据范围内的预测仍然有风险
- 能从非常小的训练集中推广

输入项变换

NeuralTools 可在训练前对数值变量进行缩放，从而使每个变量的值大致在相同的范围内。这样是为了在训练初期平衡变量对网络输出项产生的影响。如果某个变量对作出正确的预测不重要，会降低输入项与第一个隐含层神经元的连接权重，从而将这种情况在训练中反映出来。但是，如果这个不重要的变量碰巧比其他变量的数量级高，则其权重需降低更多，以补偿其过大的值。

对于每个变量，缩放采用平均值和标准偏差，对整个训练集计算。从每个值中减去平均值，然后再除以标准偏差。检验已训练的网络时，或者使用该网络进行预测时，使用相同的缩放参数。

范畴/符号数据不能在以数值作为输入项的神经网络中直接使用。所以，每个范畴自变量以一组网络输入数值代表，每个范畴一组数值。使用“n之一”转换法。例如，请看以下一组训练案例：

年龄	州	贷款额	因变量：还款情况
41	NY	4000	及时
32	CT	7000	延迟
54	NJ	6000	及时
37	NY	5000	默认

这些数据按以下方式提供给网络：

年龄	州=CT	州=NJ	州=NY	贷款额	因变量：还款情况
41	0	0	1	4000	及时
32	1	0	0	7000	延迟
54	0	1	0	6000	及时
37	0	0	1	5000	默认

推荐读物

以下文章提供了 NeuralTools 所用神经网络的更多背景信息：

Bishop, Christopher M., *Neural Networks for Pattern Recognition*, Oxford, 1995.

Masters, Timothy, *Advanced Algorithms for Neural Networks*, Wiley, 1995.

Reed, Russell D., Robert J. Marks, *Neural Smithing*, MIT, 1999.

索引

E

Evolver · 26

G

GRN 结构 · 76

M

MLF 网络与 PN/GRN 网络的比较 · 79
MLF 结构 · 71

P

Palisade Corporation · 5
PNN 结构 · 77

S

Solver · 26
StatTools · 26

三划

工具栏：NeuralTools · 31
广义回归神经网络 · 19, 44, 75

四划

专业版 · iii
分类问题 · ii
分类矩阵 · 50

五划

对话框中的图标 · 33
训练 · 15
训练 MLF 网络 · 72
训练报表 · 48
训练进度 · 47
训练命令 · 40

六划

多层前馈网络 · 44, 71
多范围数据集 · 36
安装说明 · 7
自动检验 · 42
防止过度训练 · 73

七划

均方根误差 · 50
应用程序设置命令 · 61
系统要求 · 6
运行时间 · 45

八划

图标
　桌面 · 7
图标：NeuralTools · 31
实时预测 · 24, 59
详细报表中的快速摘要 · 57

九划

卸载 NeuralTools · 7

八划

变量匹配 · 51
变量类型 · 38

九划

标签变量 · 38
残差直方图 · 50
神经网络与统计方法 · 69
神经网络管理器命令 · 64

十划

缺失值 · 67
缺失数据实用工具命令 · 66
预测 · 15, 23
预测命令 · 58
预测预览 · 59

十一划

检验 · 15
检验网络 · 22
检验报表 · 22, 53
检验命令 · 51
检验预览 · 53
综合进行训练、检验和预测 · 18

十三划

数字问题 · ii
数据集 · 16
数据集和变量 · 35
数据集和变量的容量 · 39
数据集管理器 · 17
数据集管理器对话框 · 35
数据集管理器命令 · 35
概率神经网络 · 44, 77
输入项变换 · 81